WILEY

Second Edition

SIGNALING IN TELECOMMUNICATION NETWORKS

JOHN G. VAN BOSSE • FABRIZIO U. DEVETAK

Wiley Series in Telecommunications and Signal Processing • John G. Proakis, Series Editor

Signaling in Telecommunication Networks

Second Edition

John G. van Bosse
Fabrizio U. Devetak

BICENTENNIAL 1807 WILEY 2007 BICENTENNIAL

WILEY-INTERSCIENCE
A JOHN WILEY & SONS, INC., PUBLICATION

Signaling in Telecommunication Networks

The Wiley Bicentennial–Knowledge for Generations

Each generation has its unique needs and aspirations. When Charles Wiley first opened his small printing shop in lower Manhattan in 1807, it was a generation of boundless potential searching for an identity. And we were there, helping to define a new American literary tradition. Over half a century later, in the midst of the Second Industrial Revolution, it was a generation focused on building the future. Once again, we were there, supplying the critical scientific, technical, and engineering knowledge that helped frame the world. Throughout the 20th Century, and into the new millennium, nations began to reach out beyond their own borders and a new international community was born. Wiley was there, expanding its operations around the world to enable a global exchange of ideas, opinions, and know-how.

For 200 years, Wiley has been an integral part of each generation's journey, enabling the flow of information and understanding necessary to meet their needs and fulfill their aspirations. Today, bold new technologies are changing the way we live and learn. Wiley will be there, providing you the must-have knowledge you need to imagine new worlds, new possibilities, and new opportunities.

Generations come and go, but you can always count on Wiley to provide you the knowledge you need, when and where you need it!

William J. Pesce
President and Chief Executive Officer

Peter Booth Wiley
Chairman of the Board

Signaling in Telecommunication Networks

Second Edition

John G. van Bosse
Fabrizio U. Devetak

BICENTENNIAL 1807 WILEY 2007 BICENTENNIAL

WILEY-INTERSCIENCE
A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey
Published simultaneously in Canada

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data is Available.

ISBN-13 978-0-471-66288-4
ISBN-10 0-471-66288-7

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

For
Mieps and Elizabeth

CONTENTS

PREFACE TO THE SECOND EDITION

The first edition of this book covered signaling for telecommunications through the early 1990s. Telecommunication technology has continued its remarkable progress in the years that followed. New technologies and services have come into use, and are supported by a number of new or expanded signaling systems. The aim of this second edition is to cover some of the most important of these new technologies and their signaling. Six new chapters, briefly outlined below, have been added for that purpose.

Access systems (Chapter 8) have become a part of the local network architecture. Several access systems (AS) surround a local exchange, each one serving the analog and digital lines of a group of nearby subscribers. Such systems, once called 'remote line concentrators', used to have proprietary interfaces to the local exchanges. Interfaces and signaling have now been standardized, allowing a telecom to purchase access and switching hardware equipment from different product suppliers.

In *code-division multiple access* (CDMA) wireless systems (Chapter 13) all traffic channels in a cell are carried in two common (forward and reverse) frequency channels. Individual traffic channels can be separated because the data stream from the user is encoded at the sending end with a special bit sequence that identifies the channel. To recover this information, the receiving end decodes the sender's data stream with the same sequence. Regardless of the number of cells, a CDMA system thus needs only one pair of frequency bands. This is an important advantage over AMPS and TDMA systems, which manage interference by assigning different RF bands (up to seven pairs) to adjacent cells.

INAP (Chapter 18) is the ITU-T-sponsored architecture and protocol for remote operations, equivalent to, but much more ambitious and complex than, the Advanced Intelligent Network described in Chapter 17. INAP standards are based on a sophisticated multi-level abstract view of features and services.

The first edition of this book did not cover networks for data communication, which made their appearance around the 1970s. *Data communication*

(Chapter 20) has expanded tremendously over the last 30 years, reflecting the enormous increase in the availability of computers and in the number of applications that require distributed data processing. A data message consists of a sequence of 'packets', short data bursts that have two parts: the 'header', which contains signaling information used by intermediate nodes to route the packet to its destination, and a second part that carries user information. Among the most important innovations in communications is the convergence of voice and data in packet networks, popularly known as *Voice over IP* (VoIP). The door to convergence was opened in the 1960s with the introduction of digital voice transmission. Digitally coded speech can be transported in a data network as a sequence of packets, each one carrying a number of speech samples. This technology is revolutionizing telecommunications and the driving factors for that are discussed in Chapter 20.

The most popular and widely deployed signaling protocols for VoIP, such as H.323, SIP, H.248, and BICC are discussed in Chapter 21.

ATM, or *asynchronous transfer mode* (Chapter 22), closes the book with a discussion of a packet technology that supports broadband multimedia communication, and facilitates the interworking of packet-switched and circuit- switched networks. Although widely deployed in carrier backbone networks and in ADSL access networks, ATM appears to have lost momentum to packet technology based on TCP/IP protocols.

Chapters 11 (*ISUP*), 14 (*Introduction to Transactions*), and 16 (*TCAP*) have been rewritten to cover the newest developments in ITU-T standards. Chapters 1 through 7 have also been updated to reflect current standards.

We would like to extend our grateful thanks to former colleagues from Lucent Technologies who were invaluable in providing advice and information, as well as in reviewing parts of the manuscript: Donald W Brown, James R Davis, Thomas S Hornbach, Konstantin Livanos, Alan J. Mindlin, Leon J Peeters, and Makoto Yoshida.

We cannot close this preface without also thanking our wives, Maria and Elizabeth, without whose patience, understanding, and loving support this second edition could not have been written.

1

INTRODUCTION TO TELECOMMUNICATIONS

There are two types of communication networks: *circuit-switched* networks and *packet-switched* networks. In circuit-switched networks a dedicated physical (digital or analog) circuit between the calling and called party is set up at the start of a call and released when the call has ended. Traditional telephone networks are circuit-switched networks and collectively form the switched-circuit network (SCN). Today these networks are used for speech and other purposes, such as facsimile, and are usually referred to as *telecommunication* networks.

Initially, all communication networks were circuit-switched networks. *Data networks*, consisting of a number of nodes connected by digital links, made their appearance around 1970. In these networks, a call (or *session*) consists of a series of short data bursts (packets) followed by relatively long silent intervals. A physical circuit therefore does not have to be dedicated to a single data call but can be shared by several simultaneous calls. The Internet is an example of a data network.

The terms “telecommunication network” and “data network” usually imply circuit-mode and packet-mode, respectively. However, advances in packet technology are making possible voice communication in data networks, in what is called *convergence* of voice and data. The long-term trend is toward packet communication for voice, video, and data, so the word “telecommunication” is also used sometimes to denote converged networks.

This book is about signaling in communication networks. The first nineteen chapters are dedicated to signaling in telecommunication networks, with “telecommunication” used in the traditional sense. The last three chapters are dedicated to signaling in packet networks, with focus on the convergence of voice and data.

To understand signaling it is necessary to be familiar with some basic telecommunication concepts and terms. This chapter presents an overview of telecommunication

Signaling in Telecommunication Networks, Second Edition, by John G. van Bosse and Fabrizio U. Devetak

networks (in the SCN sense). It is intended as an introduction and sets the stage for later chapters.

1.1 TELECOMMUNICATION NETWORKS

1.1.1 Introduction

Figure 1.1-1 shows a small part of a telecommunication network. It consists of *exchanges, trunks*, and *subscriber lines*. Trunks are circuits between exchanges, and the group of trunks between a pair of exchanges is known as a *trunk group* (TG). Subscriber lines (SLs) are circuits between a subscriber S and the *local* exchange (A, B, C). Exchanges D and E do not have subscriber lines and are known as *intermediate*, *tandem*, *toll*, or *transit* exchanges.

Calls. A call requires a communication circuit (connection) between two subscribers. Figure 1.1-2 shows a number of connections in the network of Fig. 1.1-1 that involve subscriber S_p. In Fig. 1.1-2(*a*), S_p is on a call with S_q who is attached to the same exchange. Calls of this type are known as *intraexchange* calls. The circuit for the call consists of the subscriber lines SL_p and SL_q and a temporary path in exchange A. Cases (*b*) and (*c*) are calls between S_p and subscribers attached to other local exchanges (*interexchange* calls). The circuit in case (*b*) consists of

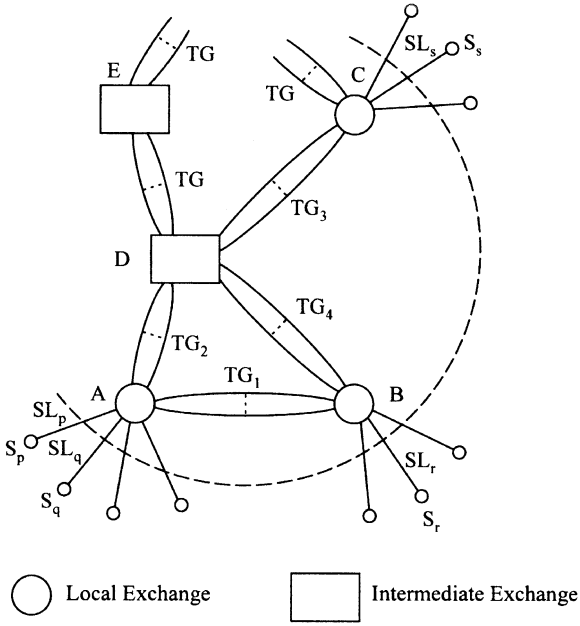


Figure 1.1-1. Partial view of a telecommunication network.

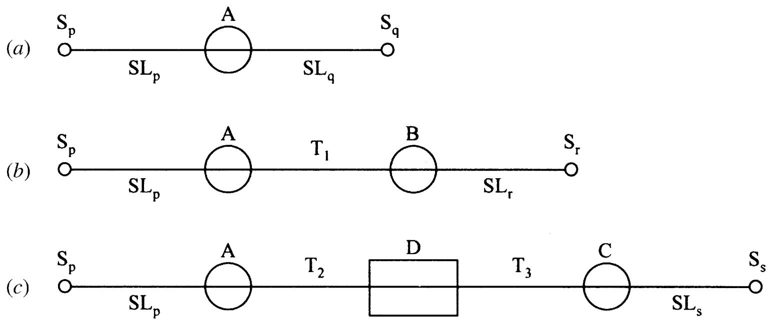


Figure 1.1-2. Connections involving subscriber S_p.

SL_p, a temporary path across exchange A, trunk T_1, a temporary path across exchange B, and SL_r. The connections of Fig. 1.1-2 are set up (switched "on") at the start of a call and released (switched "off") when the call ends.

Setup and Release. The setup and release of connections in telecommunication networks are triggered by *signals*. Starting and ending a call involve signaling between the subscribers and their local exchanges and, for interexchange calls, signaling between the exchanges along the connection.

Figure 1.1-3 shows the signaling for the setup of the connection of Fig 1.1-2(*b*). Subscriber S_p sends a request-for-service signal to exchange A (by lifting the handset of a telephone) and then signals the digits of the telephone number of S_r (with the dial or keyset of the telephone).

From the received number, exchange A determines that S_r is served by exchange B, and that the call is to be routed out on a trunk in group TG_1 (Fig. 1.1-1). It then searches for an idle trunk in this group and finds trunk T_1. Exchange A now seizes the trunk and sends a seizure signal, followed by signals that represent digits of the called number, to exchange B. It then sets up a path between SL_p and T_1.

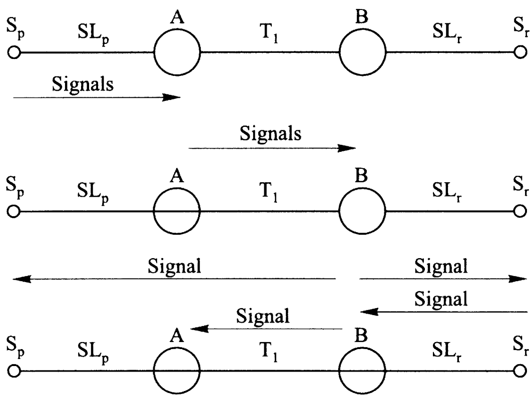


Figure 1.1-3. Setup of a connection.

When exchange B receives the seizure signal and the called number, it checks whether S_r is idle. If this is the case, it sends a ringing signal on SL_r, and a ringing-tone signal on T_1, to inform S_p. When S_r lifts the handset of the telephone, an answer signal is sent to exchange B, which then stops the ringing signal and ringing tone, sets up a path between T_1 and SL_r, and signals to exchange A that the call has been answered.

The connection is now complete and allows speech or other communications between the subscribers. At the end of the call, another signaling sequence takes place to release the connection.

One-Way and Bothway Trunk Groups. In Fig. 1.1-1, there is at most one trunk group between two exchanges. Let us consider the group TG_1. The network should allow calls originating at A with destination B and calls originating at B with destination A. Therefore, both exchanges are allowed to seize trunks in TG_1. A trunk group whose trunks can be seized by the exchanges at both ends is known as a *bothway* trunk group [1,2].

A pair of exchanges can also be interconnected by two *one-way* trunk groups. The trunks in one-way groups can be seized by one exchange only. For example, exchanges A and B could be interconnected by two one-way trunk groups TG_{1A} and TG_{1B}, whose trunks can be seized by A and B, respectively.

Both arrangements are used in actual networks. Two-way groups have an economic advantage because, for a given traffic intensity, the number of trunks of a bothway trunk group can be smaller than the total number of trunks in the one-way groups.

In bothway groups, it can happen that the exchanges at both ends of a trunk group seize the same trunk at the same time (double seizure). There are several alternatives to deal with a double seizure. For example, it can be arranged that one exchange continues the setup, and the other exchange backs off (tries to. seize another trunk for its call). The signaling on bothway trunks includes provisions to alert the exchanges when a double seizure occurs.

1.1.2 Networks

In everyday life, we think of "the" telecommunication network that allows us to speak, or send faxes and other data, to just about anybody in the world. In fact, the telecommunication network is an aggregation of interconnected networks of several types.

Networks can be classified as shown in Fig. 1.1-4. In the first place, the (global) telecommunication network consists of national networks and the international network. In turn, a national network is a combination of public and private networks. Public networks are for general use; private networks can be used only by employees of the organization (an airline company, the U.S. government, etc.) that owns the network. A public network consists of a "fixed" network and a number of "cellular mobile" networks. In the United States, the fixed public network—known as the *public switched telecommunication network* (PSTN)—consists of about 150

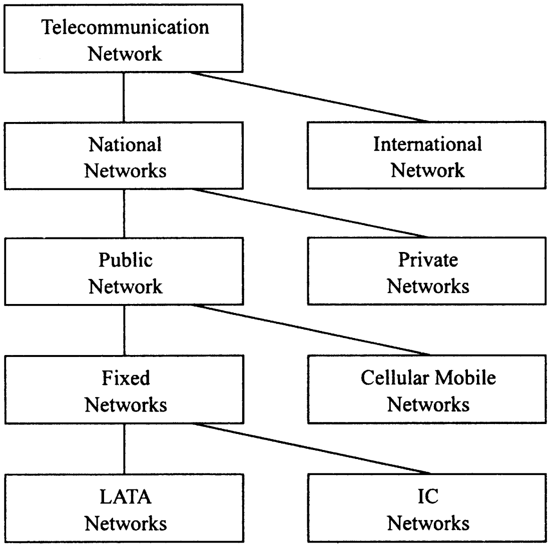


Figure 1.1-4. Networks.

LATA (*local access and transport area*) networks (the network of Fig. 1.1-1 is a LATA network), interconnected by networks that are known as IC (*interexchange carrier*), or long-distance, networks.

We now examine the interconnections of these networks. LATA and IC networks are interconnected by internetwork trunk groups—see Fig. 1.1-5. Some *local exchanges* (A) have a direct trunk group to an exchange of an IC, other exchanges (B, C, D, E) have access to the IC network via an intermediate (*tandem*) exchange in their respective LATAs.

A cellular network has one or more *mobile switching centers* (MSCs)—see Fig. 1.1-6. Each MSC is connected by an internetwork trunk group to a nearby tandem exchange T of a fixed (LATA) network. When a mobile station is making a call, it uses a radio channel of a nearby MSC.

Private Branch Exchange (PBXs). These are exchanges owned by government agencies, businesses, and so on and located in buildings that belong to these organizations. A PBX enables the employees in a building to call each other and to make and receive calls from subscribers served by the public network. A PBX is connected by an *access line group* (ALG) to a nearby local exchange (Fig. 1.1-7).

An organization with PBXs in several cities can establish a *private network* that consists of the PBXs and a number of *tie trunk groups* (TTG) between the public local exchanges to which their PBXs are attached. A TTG is a "private" group that is leased by the LATA operator and is dedicated to private-network calls. In Fig. 1.1-7, the connection for a call between public branch exchanges X and Y uses a trunk of TTG_1 and is switched in the public local exchanges A and B.

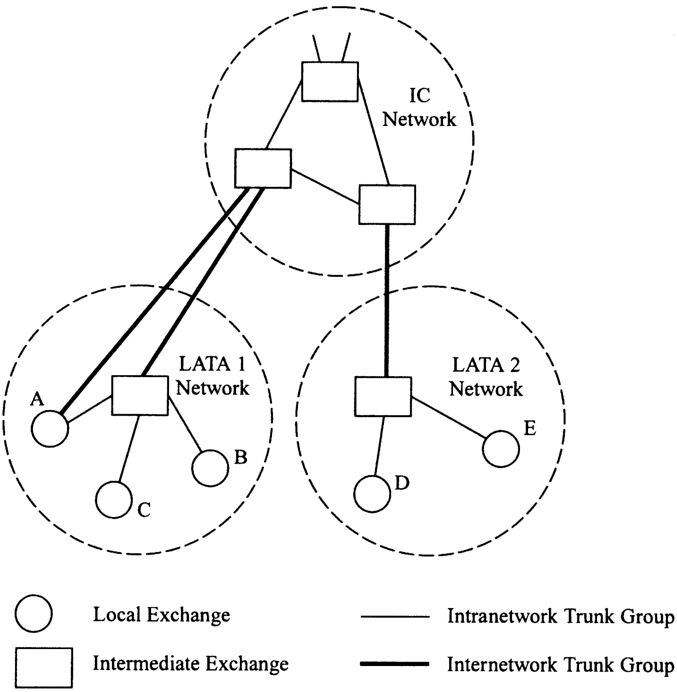


Figure 1.1-5. Interconnection of LATA and IC networks.

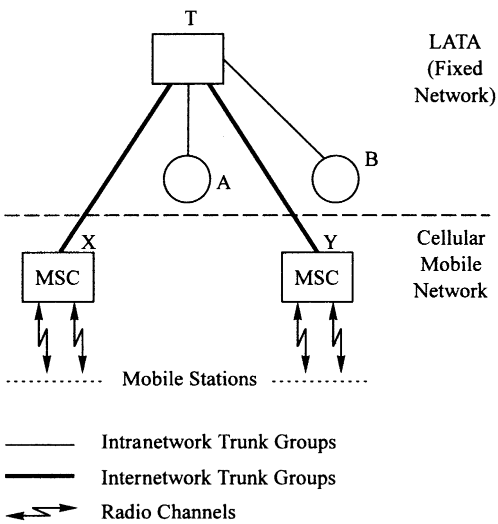


Figure 1.1-6. Interconnection of fixed and mobile networks. MSC: mobile switching center.

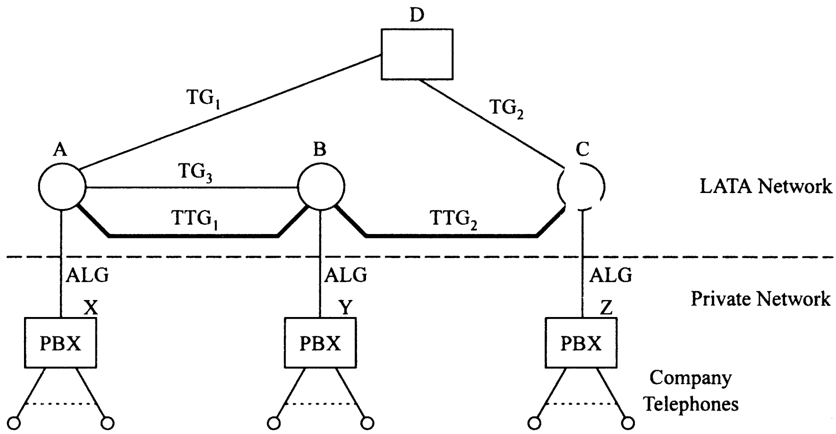


Figure 1.1-7. Interconnection of private network and a LATA network. ALG, access line group; TG, trunk group (public); TTG, tie trunk group (private); PBX, private branch exchange.

Today there are also *virtual private networks* (VPNs). They appear to a business as a private network but use the trunks of the public networks.

International Calls. Figure 1.1-8 shows the interconnection of long-distance (IC) networks in different countries. For a call from country A to country C, an IC network in country A routes the connection to an *international switching center* (ISC). An ISC has national trunk groups to exchanges of its IC and international trunk groups to ISCs in foreign countries.

The term "international network" refers to the combination of the ISCs and their interconnecting trunk groups.

1.1.3 Telecoms

We shall use the term *telecom* to denote a company that owns and operates a public telecommunication network. Until recently, the telecoms in most countries were government-owned monopolies that operated an entire national network. In recent years, a number of countries have started to privatize their telecoms and to allow competition by newly formed telecoms.

The networks in the United States are operated by investor-owned telecoms. Until 1984, the Bell System was the largest telecom, operating practically the entire long-distance network and many—but by no means all—local networks. Independent (non-Bell) telecoms, such as GTE, United Telecoms, and a host of smaller companies, operated the other local networks.

The division of the U.S. public network into *local access and transport areas* (LATAs) and *interexchange carrier* (IXC or IC) networks took place in 1984, due to government actions that broke up the Bell System [1]. As a result, competing IXCs started offering long-distance and international service, while *local exchange*

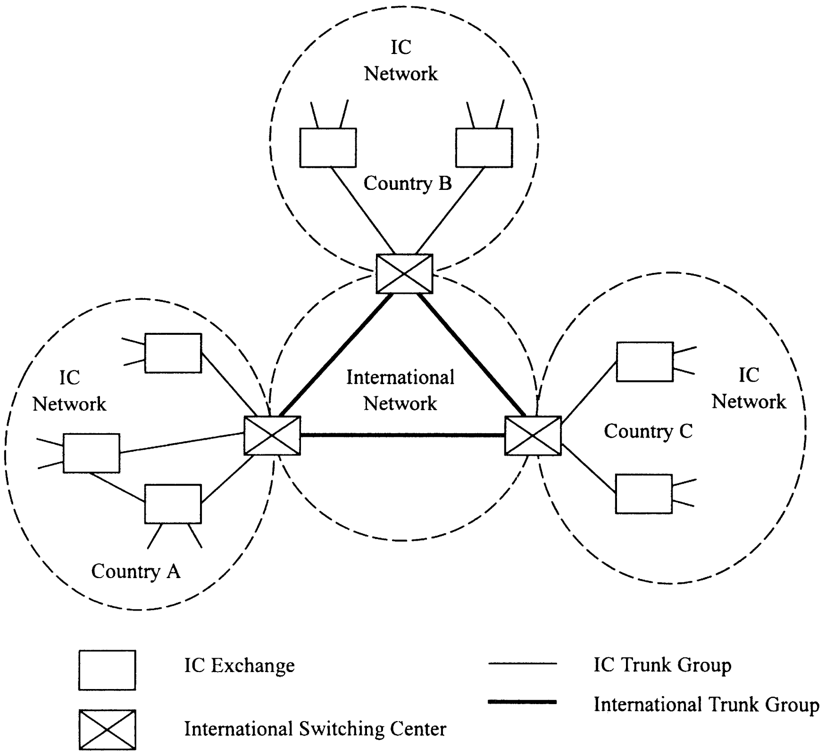


Figure 1.1-8. Interconnections between national IC networks and the international network.

carriers (LECs) provided service on a regional basis, in LATAs. A typical LEC owned a number of adjacent predivestiture local networks. Competition for local service was allowed but was slow in developing because of high entry barriers. The latest development is that IXCs and LECs are allowed to provide both local and long-distance service. In addition, new wireline, cable, and wireless companies can provide competing telephone service. All that has resulted in mergers between LECs, IXCs, cable companies, and wireless companies.

1.1.4 Synonyms

Since telecommunication terms originated rather independently in different countries, technical literature in English still uses different terms for the same concepts, depending on whether the authors are from the United States, from the United Kingdom, or from other English-speaking countries, or documents are translations from other languages. Some frequently used synonyms are listed below:

Subscriber, customer, user

Subscriber line, line, loop

Local exchange, local office, central office, end office

Intermediate exchange, tandem exchange, toll exchange, transit exchange

International switching center, gateway, international exchange

Trunk, junction, circuit

Telecom, administration, carrier, operating company, telephone company, telco, service provider

Exchange, switch

Switchblock, switch fabric

1.2 NUMBERING PLANS

This section explores the formats of the numbers (sometimes called addresses) that identify the subscribers of telecommunication networks.

Subscriber Numbers (Directory Numbers). The geographical area of a nation is divided into several *numbering areas*, and subscriber numbers (SNs) identify subscriber lines within a particular numbering area. A SN consists of an *exchange code* (EC) that identifies an exchange within a numbering area, followed by a *line number* (LN):

$$SN = EC\text{-}LN$$

National Numbers. Within a country, a subscriber is identified by a national number (NN), consisting of an *area code* (AC), which identifies the numbering area, followed by a subscriber number:

$$NN = AC\text{-}SN = AC\text{-}EC\text{-}LN$$

International Numbers. Worldwide, a subscriber is known by an *international number* (IN) that consists of a country code (CC), followed by a national number:

$$IN = CC\text{-}NN$$

The generic format for the international numbering plan is specified by ITU-T in Rec. E.164 [2]. Three types of numbering schemes are supported by E.164, all with the CC component having up to three digits and the total IN number having a maximum of 15 digits:

1. Numbering plan for geographic areas (subscriber numbers)
2. Numbering plan for global services (e.g., Freephone numbers)
3. Numbering plan for networks (other than the telephone network)

For item 1 the CC may have one, two, or three digits and the NN breakdown is left open for national definitions. For global services and for networks, the CC must be three digits long. For networks, the NN is divided into an *identification code* (IC) of one to four digits and a subscriber number.

When subscriber S_1 calls a subscriber located in the same numbering area, the number dialed is a SN. If the called subscriber lives in the same country but in a different area, S_1 has to dial a NN. If the called party lives in another country, S_1 needs to dial an IN. To allow the local exchange to interpret what is being dialed, prefixes may need to be prepended to the numbers of a numbering plan. Prefixes are part of the *dialing plan*, described in Section 3.7.

National numbering plans define the formats of subscriber and national numbers. Most countries have their own numbering plans. However, the United States, Canada, and a number of Caribbean countries are covered by a common plan that was introduced in the mid-1940s.

1.2.1 North American Numbering Plan (NANP)

The North American territory [1] is divided into *numbering plan areas* (NPAs), which are identified by *three-digit numbering plan area codes*, most often called simply *area codes*, AC(3).

Each area covers a state, or part of a state, but never crosses a state boundary. Lightly populated states (Alaska, New Mexico, etc.) have one NPA, while more heavily populated states (Illinois, New York, etc.) are divided into several NPAs. The territory of a NPA is not identical to the service area of a LATA (LATA boundaries were established much later, after the breakup of the Bell System). Some (but not all) of the AC(3) numbering space in the 8XX and 9XX ranges is used for services where destinations are not directly linked to a geographical area ("800 service" and "900 service"). The 700 NPA is also available for individual carriers who want to introduce new services.

A subscriber number has seven digits: a three-digit exchange code, which defines an exchange within a NPA, followed by a four-digit line number:

$$\text{SN(7)} = \text{EC(3)-LN(4)}$$

Each EC can cover maximally 10,000 line numbers. Since local exchanges can serve up to some 100,000 subscribers, more than one exchange code may have to be assigned to a particular exchange. For example, in a particular NPA the subscriber numbers 357-XXXX, 420-XXXX, and 654-XXXX might be served by the same local exchange.

National numbers consists of ten digits: a three-digit area code AC(3), followed by a seven-digit subscriber number:

$$\text{NN(10)} = \text{AC(3)-SN(7)} = \text{AC(3)-EC(3)-LN(4)}$$

The NANP (for national numbers) is an example of a closed (uniform) numbering plan. In these plans, the lengths of all subscriber numbers, and of all national numbers, are constant.

Originally, the format of AC(3) was NXX with $N = 2$ or 3 and $X = 0{-}9$, and the format of EC(3) was NXX with $N = 4{-}9$ and $X = 0{-}9$. This allowed up to 160 area codes and up to 640 exchange codes. The increase in telephones in later years required a larger number of area codes, which led to a change resulting in the current format for both AC(3) and EC(3):

$$NXX \quad (N = 2{-}9 \text{ and } X = 0{-}9)$$

that has a capacity of 800 area codes and 800 exchange codes.

To support competition by new long-distance carriers after the 1984 divestiture, a four-digit *carrier identification code* CIC(4) may be inserted before the called party number to route the call through the long-distance carrier identified by the CIC(4) instead of going through the presubscribed IXC.

Prefixes. In order to expedite the setup of a connection, there are situations in which a subscriber has to dial one or more "prefix" digits ahead of the called party address. For example, since AC(3) and EC(3) codes use the same format, an exchange would have to perform a time-out after receiving the seventh digit to distinguish a subscriber number from a national number. Therefore, subscribers must dial the prefix "1" to indicate that the called address is a NN, while prefix "011" indicates that the address is an international number. Similarly, the prefix "101" must he dialed when a carrier identification code is inserted before the called party number. More details on prefixes are found in Section 3.7.1.

1.2.2 Other National Numbering Plans

Some countries have *open* numbering plans, in which subscriber numbers and area codes (sometimes called *trunk* or *city* codes) are not of fixed length. In these plans, the numbering areas usually have comparable geographical sizes. Heavily populated areas need subscriber numbers with six or seven digits, while four or five digits are sufficient in lightly populated areas. To limit the differences in length of national numbers, the area codes for areas with long subscriber numbers are usually shorter than those for areas with short subscriber numbers. An example of national numbers in an open numbering plan is shown below:

$$\text{NN}(8) = \text{AC}(2)\text{-SN}(6)$$
$$\text{NN}(9) = \text{AC}(2)\text{-SN}(7)$$
$$\text{NN}(8) = \text{AC}(4)\text{-SN}(4)$$
$$\text{NN}(9) = \text{AC}(4)\text{-SN}(5)$$

In order to allow exchange to interpret national numbers in this plan, the two initial digits of a four-digit area code cannot be the same as those of a two-digit area code. For example, the two-digit area code 70 precludes the use of four-digit area codes 70XX.

1.2.3 Country Codes

The country codes have been established by ITU-T [2] and consist of one, two or three digits. The first digit indicates the world zone in which the called party is located:

World Zone

1: North America
2: Africa
3: Europe
4: Europe
5: Latin America
6: Australia and Southern Pacific Region
7: Former Soviet Union
8: China and Northern Pacific Region
9: Middle East

Country codes starting with 1 and 7 are one-digit codes and represent, respectively, North America and the former Soviet Union. Country codes starting with 2 through 9 can have two- or three-digit codes, and the combinations of the first and second digit determine which is the case. For instance, in world zone 3, all combinations except 35 are two-digit codes, as shown by the following examples:

31	The Netherlands
354	Iceland
359	Bulgaria

These rules enable exchanges to separate the country code from the national number in a received international number.

Country code 1 represents the United States, Canada, and a number of Caribbean countries. Most area codes represent areas in the United States. Other codes represent areas in Canada and individual Caribbean nations.

1.2.4 Digit Deletion

In Fig. 1.2-1, calling party S_1 and called party S_2 are located in different NPAs of the United States. S_1 therefore dials the national number (NN) of S_2. As a general rule, exchanges send subscriber numbers to exchanges that are in the NPA of the destination exchange, and national numbers to exchanges outside the destination

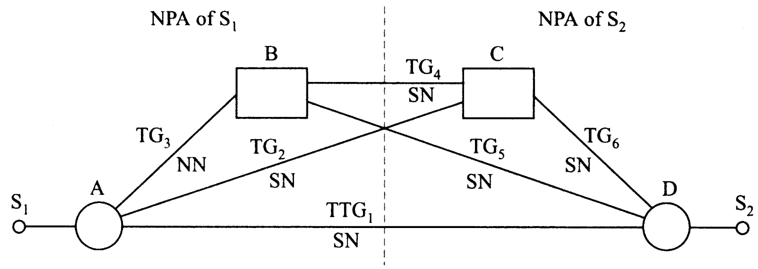


Figure 1.2-1. Called numbers. SN, subscriber number; NN, national number.

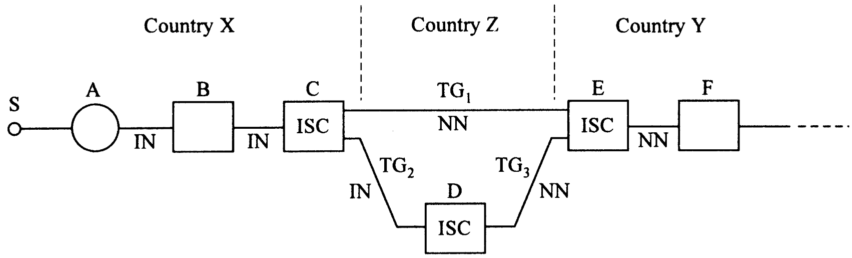


Figure 1.2-2. Called number formats on international calls. NN, national number; IN, international number.

NPA. In Fig. 1.2-1, exchange C is in the NPA of D, and exchanges A and B are not. Exchange A has received the NN of S_2. If A routes the call on a trunk of TG_1 or TG_2, it deletes the area code from the received number and sends the SN of S_2. However, if A routes the call on TG_3, it has to send the NN of S_2. Exchanges Band C always send the SN.

A similar digit deletion occurs on international calls. Figure 1.2-2 shows a call from subscriber S in country X to a subscriber in country Y. An ISC sends a national or international number, depending on whether it routes the call on a direct trunk on an ISC in the destination country, or on a trunk to an ISC in an intermediate country.

1.3 DIGIT ANALYSIS AND ROUTING

1.3.1 Destinations and Digit Analysis

Connections for interexchange calls are set up along paths that have been predetermined by the network operator. A *route* is a path to a particular *destination*. An exchange determines the call destination by analyzing the called number and then selects an outgoing trunk in a route to the destination.

We need to distinguish two destination types. The *final destination* (FDEST) of a call is the local exchange that serves the called party. An *intermediate destination*

(IDEST) is an exchange where the call path enters another network, on its way to the final destination. For a connection, the destination at the exchanges of the local network serving the called party is a FDEST. The destinations at exchanges in the other networks are IDESTs.

As an example, take a call from a calling party in local network $LATA_1$ to a called party in $LATA_2$. The interexchange carrier designated by the calling party is IC_D. In the exchanges of $LATA_1$, the call has an IDEST, namely, an exchange in the network of IC_D, predetermined by the telecoms of the $LATA_1$ and IC_D networks. In the IC_D exchanges, the call also has an IDEST: an exchange in $LATA_2$ predetermined by the telecoms of IC_D and $LATA_2$. In the exchanges of $LATA_2$, the call has final destination.

Digit analysis is the process that produces a FDEST or an IDEST from the called subscriber number (EC-LN), national number (AC-EC-LN), or international number (CC-AC-EC-LN).

In LATA exchanges, calls with subscriber and national numbers can have IDEST or FDEST destinations. Calls with international called numbers always have an IDEST. In IC exchanges, all calls have IDEST destinations. In calls with national called numbers, the IDEST is an exchange in the LATA network determined by the combination AC-EC.

For calls with international called numbers, the IDEST depends on whether the IC exchange is an ISC. If the exchange is not an ISC, the call destination is an ISC in the IC network, determined by the country code (CC) in the number. At an ISC, the destination is an ISC in the country identified by CC.

1.3.2 Routing of Intra-LATA Calls

Intra-LATA calls are handled completely by one telecom. In these calls, the FDEST is the local exchange of the called party. We examine a few routing examples for calls from a caller on local exchange A to a called party on local exchange D.

In Fig. 1.3-1(*a*), the telecom of the LATA has specified one indirect route, consisting of trunk groups TG_1 and TG_2:

Route A-TG_1-X-TG2-D

The fact that a TG belongs to a route does not mean that the TG is dedicated to the route. For example, TG_1 can also belong to routes from A to other destinations.

In Fig. 1.3-1(*b*), the telecom has specified a set of four routes:

A-TG_3-D
A-TG_4-Y-TG_6-D
A-TG_1-X-TG_2-D
A-TG_1-X-TG_5-Y-TG_6-D

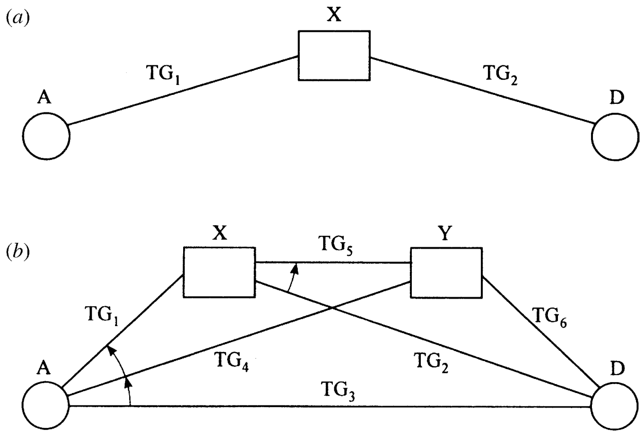


Figure 1.3-1. Routes for calls from A to D (intra-LATA calls).

As perceived by an exchange, a route to a destination is an outgoing trunk group (TG). In Fig. 1.3-1(*b*), the route set at exchange A for destination D consists of trunk groups TG_1, TG_3, and TG_4.

Each exchange has a list of routes that can be used for a destination. The lists for destination D at exchanges A, X, and Y are:

Exchange	**Routes for Destination D**
A	TG_3, TG_4, TG_1
X	TG_2, TG_5
Y	TG_6

Alternate routing is the procedure by which an exchange selects an outgoing trunk for a call when there are several routes to a destination. In this procedure, the order in which the routes are listed specifies the sequence in which an exchange checks the outgoing trunk groups for available trunks. In this example, exchange A first tries to find an available trunk in its first-choice route TG_3. If a trunk is available, A seizes the trunk. If not, it attempts to find an available trunk in its second-choice route TG_4, and so on. If none of these routes has an available trunk, exchange A aborts the setup of the call.

In alternate routing, the TGs to a destination are ordered such that the first-choice route is the most direct one (passing through the smallest number of intermediate exchanges), the second-choice route is the most direct one among the remaining routes, and so on. In Fig. 1.3-1(*b*), the arrows indicate the selection sequences at exchanges A and X.

1.3.3 Routing of Inter-LATA Calls

Figure 1.3-2 shows a routing example for a call originated by a subscriber attached to local exchange A of $LATA_1$, to a called party attached to local exchange Z in $LATA_2$.

In $LATA_1$, the IDEST of the call is exchange P or Q, in the IC network designated by the calling subscriber—Fig. 1.3-2(*a*). The routes for this destination at exchanges A, B, and C are:

Exchange	Routes for IDEST
A	TG_1, TG_2
B	TG_3, TG_4
C	TG_5, TG_6

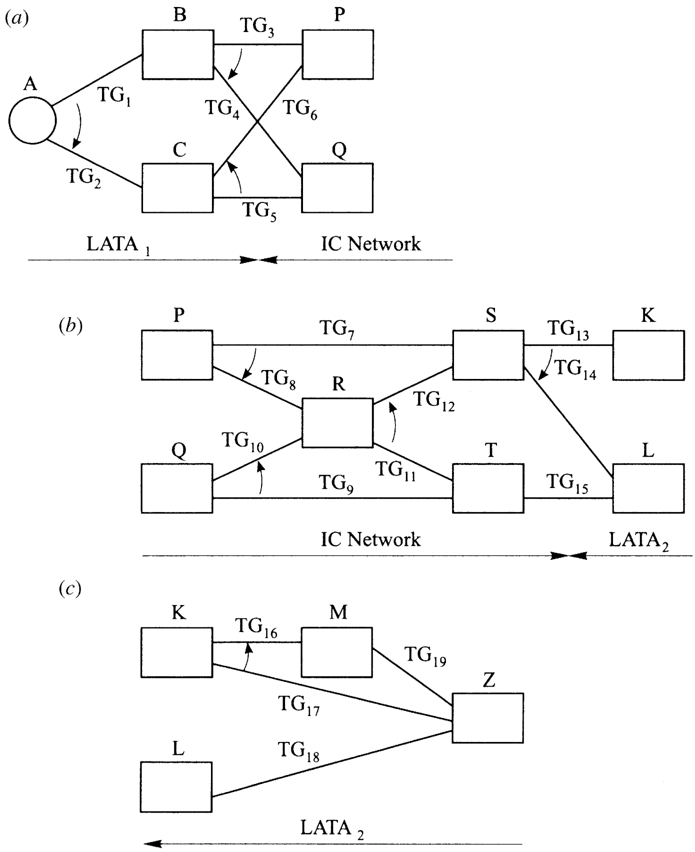


Figure 1.3-2. Routing of inter-LATA call: (*a*) routing in $LATA_1$, (*b*) routing in IC network, and (*c*) routing in $LATA_2$.

In the IC network, the IDEST for the call is exchange K or L $LATA_2$—Fig. 1.3-2(*b*)—and the routes for this destination at exchanges P, Q, R, S, and T are:

Exchange	Routes for IDEST
P	TG_7, TG_8
Q	TG_9, TG_{10}
R	TG_{11}, TG_{12}
S	TG_{13}, TG_{14}
T	TG_{15}

Finally, in $LATA_2$, the FDEST of the call is local exchange Z—Fig. 1.3-2(*c*)—and the routes at exchanges K, L, and M are:

Exchange	Routes for IDEST
K	TG_{17}, TG_{16}
L	TG_{18}
M	TG_{19}

1.3.4 Automatic Rerouting

Automatic rerouting (also called *crankback*) is a refinement of *alternate routing*. It is used in AT&T's long-distance network [3]. The procedure is illustrated with the example of Fig. 1.3-1(*b*).

Suppose that TG_3 is congested, and that A has seized trunk T in its second-choice route TG_4. The call setup arrives at exchange Y, which has only one route (TG_6) to destination D. Under alternative routing, if no trunk is available in TG_6, exchange Y abandons the setup and informs the calling party with a tone or recorded announcement.

Under automatic rerouting, exchange Y signals to exchange A that it is unable to extend the setup. A then releases trunk T and tries to route the call on its final route (TG_1). If trunks are available in TG_1 and TG_2, the connection can be set up.

Automatic rerouting depends on the ability of an exchange (Y) to signal the preceding exchange (A) that it is not able to extend the setup. We shall encounter signaling systems that have signals for this purpose.

1.4 ANALOG TRANSMISSION

Until 1960, analog transmission was the only form of transmission in telecommunication networks. Today, the telecommunication network is mostly digital, except for the subscriber lines.

This section outlines some basic aspects of the transmission of analog signals in telecommunications. In this section, the term *signal* refers to information (speech or voiceband data) exchanged between subscribers during a call (as opposed to *signaling*, which is the subject of this book).

1.4.1 Analog Circuits

An analog signal is a continuous function of time [4,5]. Telecommunication started out as telephony, in which a microphone (or transmitter, mouthpiece) produces an electrical *analog* signal, whose variations in time approximate the variations in air pressure produced by the talker's speech. A receiver (or earpiece) reconverts the electrical speech signal into air-pressure variations that are heard by the listener.

The pressure variations of acoustic speech are complex and not easily described. "Average" acoustic speech contains frequencies from 35 Hz to 10,000 Hz. Most of the speech power is concentrated between 100 Hz and 4000 Hz. For good-quality telephony, only the frequencies between 300 and 3400 Hz need to be transmitted. Analog communication channels in the network, which historically have been designed primarily for speech transmission, therefore accommodate this range of *voiceband* frequencies (Fig. 1.4-1).

Two-Wire and Four-Wire Circuits. Analog circuits can be two-wire or four-wire. Subscriber lines are two-wire circuits, consisting of a pair of insulated copper wires that transfer signals in both directions. Most analog trunks are four-wire circuits, consisting of two unidirectional two-wire circuits, one for each direction of transmission.

In the diagrams of this section, circuits are shown as in Fig. 1.4-2. The arrows indicate the directions of transmission. In general, bidirectional circuits (two- or four-wire) are shown as in (*a*), and a note indicates whether the circuit is

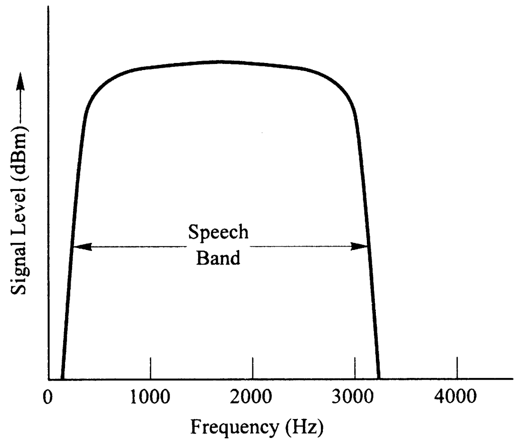


Figure 1.4-1. Frequency response of analog transmission circuits. (From R. L. Freeman, *Telecommunication System Engineering*, 2nd ed., Wiley, 1992.)

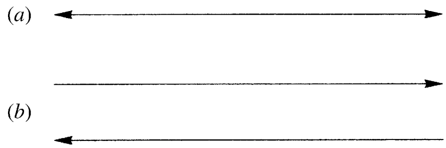


Figure 1.4-2. Circuit representations: (*a*) bidirectional two-wire or four-wire circuit and (*b*) two unidirectional two-wire circuits forming a bidirectional four-wire circuit.

two-wire or four-wire. When discussing the two unidirectional circuits of a four-wire circuit, representation (*b*) is used.

Data Transmission on Analog Circuits. In the 1960s, subscribers also began to use the telephone network for transmission of digital data, and the network has become a *telecommunication* network. Digital data are converted by *modems* into a form that fits within the 300–3400 Hz band of analog circuits. There are several modem types. *Frequency-shift keying* (FSK) modems convert the zeros and ones of the digital bit stream into two voiceband frequencies, for example, 1300 and 1700 Hz. With FSK, data can be sent at speeds of 600 or 1200 bits/second. The signal produced by a *differential phase-shift keying* (DPSK) modem is a single frequency with phase shifts. In the widely used V.26 modem [6], the frequency is 1800 Hz, and the phase shifts occur at a rate of 1200 shifts/second. The phase shifts can have four magnitudes, each of which represents the values of two consecutive bits in the digital signal. The modem thus transfers 2400 bits/second. More recently developed modems have transfer rates of up to 56 kb/second.

1.4.2 Analog Subscriber Lines [4,5]

Figure 1.4-3 shows a connection between two subscribers served by an analog local exchange. The subscriber lines (SLs) and the path (P) across the exchange are two-wire circuits.

The power of an electrical signal decreases as it propagates along the circuit. This attenuation becomes more severe with increasing circuit length.

The characteristics of microphones and receivers are such that a listener receives a sufficiently strong acoustical signal when at least 1% of the electrical signal power produced by the talker's microphone reaches the listener's receiver. This corresponds to the attenuation in a circuit of about 15 miles. Most subscriber

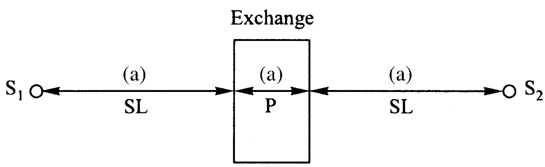


Figure 1.4-3. Circuit for intraoffice call. (a), Two-wire analog circuits.

lines are less than 4 miles long, and there are no signal-strength problems in intra-exchange calls.

1.4.3 Two-Wire Analog Trunks

Two-wire trunks are similar to subscriber lines and have similar attenuation characteristics. This limits the trunk length to about 10 miles.

1.4.4 Four-Wire Analog Trunks

Long-distance trunks require amplification to compensate the signal attenuation. Amplifiers are unidirectional devices, and this is why long-distance trunks are four-wire trunks. A four-wire circuit consists of two amplified unidirectional two-wire circuits. In Fig. 1.4-4, *hybrid* circuits (H) at both ends of the trunk convert a two-wire circuit into a four-wire circuit, and vice versa. Amplifiers (A) are located at regular intervals along the two unidirectional circuits. The unidirectional circuits at an exchange that transfer signals to and from the distant exchange are known as the send circuit (S) and the receive circuit (R).

1.4.5 Frequency-Division Multiplexing

Frequency-division multiplexing (FDM) [4,5] is a technique to carry the signals of a group of n analog four-wire trunks on a common four-wire analog transmission system. In each direction of transmission, the bandwidth of the transmission system is divided into n 300–3400 Hz channels, spaced at intervals of 4 kHz, for a total bandwidth of $n \times 4$ kHz (Fig. 1.4-5).

Figure 1.4-6 shows a group of n FDM trunks between exchanges X and Y. In this example, the trunks appear at the exchanges as two-wire circuits (a) that can transfer analog signals in the 300–3400 Hz range.

Frequency-division multiplexing (FDM) equipment is located at both exchanges. Hybrids (H) convert the two-wire circuits into four-wire circuits. For transmission in direction X $\rightarrow$ Y, multiplexer FDM-X shifts the 300–3400 Hz send channels (S) of

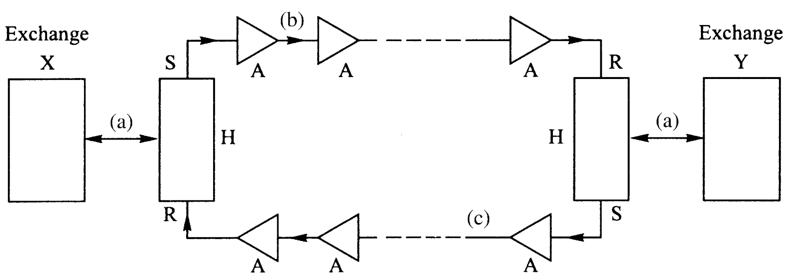


Figure 1.4-4. Amplified four-wire analog trunk. (a),Two-wire bidirectional analog circuits; (b,c), pair of unidirectional analog circuits forming a bidirectional circuit.

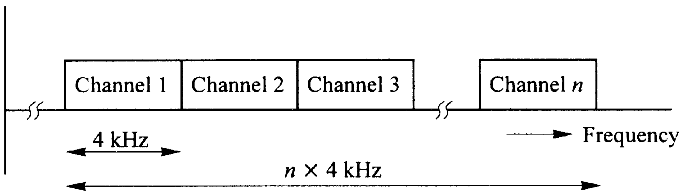


Figure 1.4-5. Frequency allocation of FDM transmission system.

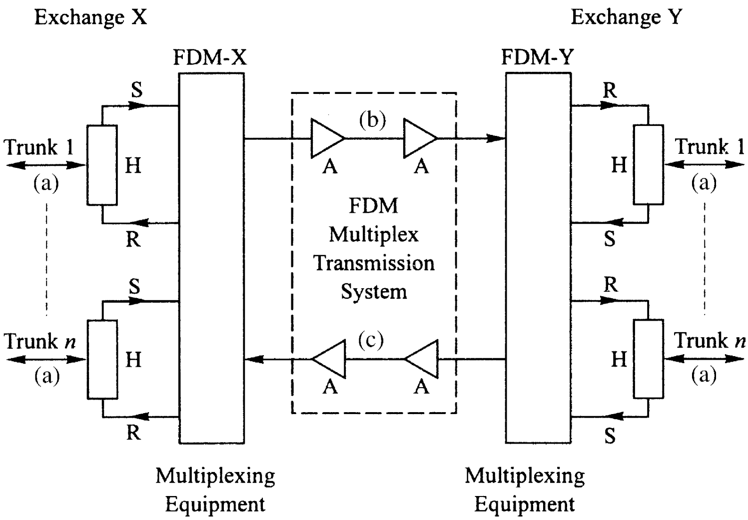


Figure 1.4-6. Multiplexed analog trunks. (a), Two-wire bidirectional circuit; (b,c), multiplexed unidirectional circuits (*n* channels), forming a four-wire FDM transmission system; A, amplifiers; S, R, two two-wire unidirectional circuits, forming a bidirectional circuit.

the trunks "up" in frequency (by different amounts) and then combines the channels on a unidirectional multiplex transmission circuit (b). FDM-Y demultiplexes the signal received on (b), by shifting the channels downward in frequency. The demultiplexed channels are the 300–3400 Hz receive channels (R) of the trunks at exchange Y.

In the other direction, the send channels at exchange Y are multiplexed by FDM-Y, combined on multiplex transmission circuit (c), and demultiplexed by FDM-X. Amplifiers A along the multiplexed circuits (b) and (c) compensate the signal attenuation.

Larger FDM multiplex systems can be formed by repeated multiplexing. This creates a FDM multiplex hierarchy, which has been standardized internationally [5]. In Fig. 1.4-7, the 60 two-wire trunks T are converted to four-wire circuits (b) by hybrids (H). The first-order FDM multiplexes convert 12 four-wire circuits into a four-wire *group* circuit (c) with 12 channels. The bandwidth of group circuits is $12 \times 4 = 48$ kHz in each direction (60–108 kHz).

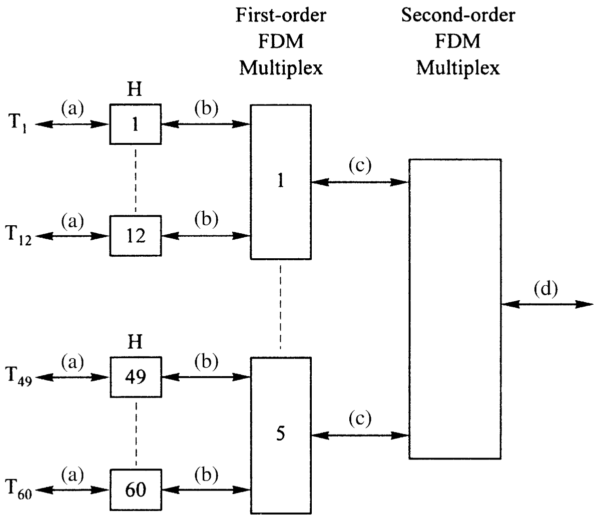


Figure 1.4-7. First- and second-order FDM multiplexes. (a), Two-wire bidirectional circuits; (b), four-wire circuits; (c), multiplexed four-wire circuits (12 channels); (d), multiplexed four-wire circuits (60 channels).

The second-order FDM multiplex combines five group circuits (c) into one four-wire *supergroup* circuit (d) with 60 channels. The bandwidth of a supergroup circuit is 5×48 kHz $= 240$ kHz in each direction (312–552 kHz).

Continuing in this way, ten supergroups can be combined into a 600 channel *mastergroup*, and six mastergroups form a 3600 channel *jumbogroup* [5].

1.4.6 FDM Transmission Systems

Group circuits can be carried on two amplified wire-pairs, with amplifiers at regular spacings. Signals of higher order multiplexes are carried on transmission systems of several types.

Cable transmission systems consist of two coaxial cables, with amplifiers at regular spacings. Cable systems are used on overland and underwater (transatlantic, transpacific) routes.

Microwave radio systems consist of a pair of unidirectional *radiofrequency* (RF) transmission links in the microwave region (2, 4, 6, 11, or 18 GHz). The output signal of a FDM multiplexer modulates the frequency of the microwave carrier. The RF signal travels in a narrow beam from the transmitting antenna to the receiving antenna.

In terrestrial systems, the microwave links are divided into several sections (hops) of, say, 20–40 miles. The transmission in each section is on a line-of-sight path. Repeater stations at the section boundaries amplify the RF signal received from one section and retransmit it to the next section. The repeater stations are land based, and terrestrial microwave systems can therefore be used on overland routes only.

Satellite microwave systems use communication satellites. These satellites are in *geosynchronous* orbits, some 22,300 miles above the equator. At this altitude, a satellite circles the earth at an angular velocity equal to the angular velocity of the earth's rotation. When observed from a point on earth, the satellite therefore appears in a fixed position, and only small and infrequent adjustments are needed to keep an antenna of a ground station aimed at the satellite. Each satellite link consists of an "uplink" from a ground station to the satellite, and a "downlink" from the satellite to another ground station. The satellite amplifies the signal received on the "uplink" and retransmits it on the "downlink."

Satellite microwave systems were introduced in the 1960s and have been used extensively on transoceanic routes. A disadvantage of satellite transmission is the long signal propagation time (in excess of 250 ms from ground station to ground station), which can cause problems in telephone conversations.

1.5 DIGITAL TRANSMISSION

Analog signals (a) can be converted to digital bit streams (d) and transmitted on digital transmission systems (Fig. 1.5-1).

Digital transmission is more robust and of better quality than analog transmission. While digital signals require more bandwidth, digital transmission systems are often less expensive than their analog counterparts, and digital transmission is rapidly replacing analog transmission in telecommunications.

1.5.1 Pulse Code Modulation

Pulse code modulation (PCM) is the earliest method for *analog-to-digital* (A/D) and *digital-to-analog* (D/A) conversions used in telecommunication networks [4,5]. The basic idea dates back to 1938 [7]. However, PCM requires a fair amount of digital signal processing and only became technically feasible after the

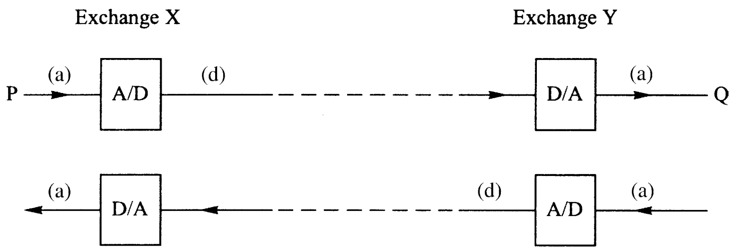


Figure 1.5-1. Digital transmission system. (a), Analog signals; (d), digital bit stream; A/D, analog-to-digital converter; D/A, digital-to-analog converter.

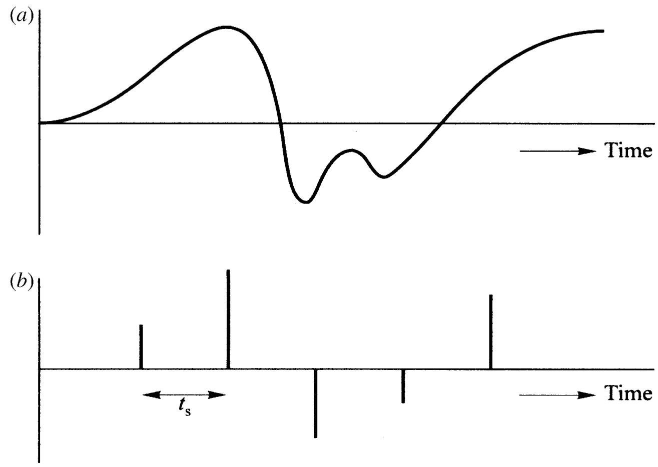


Figure 1.5-2. Sampler (*a*) input and (*b*) output.

invention of semiconductor devices. The first PCM systems were installed in the United States by the Bell System, in 1962 [1,4].

PCM is designed such that its frequency response, measured from analog input signal (P) to analog output signal (Q) of Fig. 1.5-1, is the same (300–3400 Hz) as for FDM analog transmission—see Fig. 1.4-1.

In PCM, the analog input signal (a) of an A/D converter is sampled at $t_s = 125$ microsecond intervals (8000 samples/second)—see Fig. 1.5-2. The magnitude of each sample is then converted into an *octet* (an eight-bit binary number). The digital signal (d) of Fig. 1.5-1 is a sequence of octets, transmitted at $8 \times 8 = 64$ kb/s (kilobits/second).

PCM Standards. In order to preserve the quality of low-level speech, the A/D and D/A conversions are nonlinear. Two coding rules are in existence: mu-law coding is the standard in the United States, Canada, and Taiwan, and A-law coding is the standard in the rest of the world [7,8].

1.5.2 Time-Division Multiplexing

PCM trunks are carried on four-wire *time-division multiplexing* (TDM) transmission systems [8]. Figure 1.5-3 shows a first-order pulse code modulation multiplex (PCM-X) for *m* trunks. In the example, the trunks are attached to the exchange as two-wire analog trunks (a) and are converted to four-wire analog circuits by hybrids (H).

A multiplexer PCM-X does the A/D conversions of the signals in the analog *send channels* (S) and then multiplexes the resulting octets into the outgoing bit stream (b). PCM-X also segregates the octets of the individual channels in the incoming

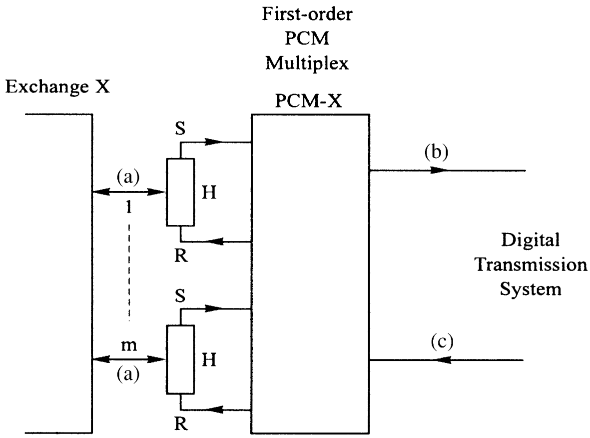


Figure 1.5-3. PCM first-order multiplex, attached to a two-wire analog exchange. (a), Analog two-wire bidirectional circuits; S, R, send and receive circuits of analog four-wire circuits; (b,c), outgoing and incoming circuits of a four-wire multiplexed digital circuit (*m* channels); H, hybrids.

bit stream (c) and does the D/A conversions that produce the voiceband analog signals in the *receive channels* (R).

T1 and E1 Digital Transmission Systems. There are two standards for first-order transmission systems. The T1 system, a development of Bell Laboratories, is used mainly in the United States [1,8]. The E1 system has been defined by the European Telecommunications Standards Institute (ETSI) and is used in most other countries [8].

In both systems, the bit stream is divided in *frames* that are transmitted at a rate of 8000 frames/second. Each frame is divided into a number of eight-bit time-division channels, known as *time slots* (TS), that contain the PCM samples of the individual trunks.

The frame format of the T1 system (known as the *DS1 format*) is shown in Fig. 1.5-4(*a*) [4,5,8]. It consists of $m = 24$ time slots (TS_1 through TS_{24}) and a framing bit (F). The F bits in successive frames form a fixed *synchronization* sequence that repeats every 12 frames. This enables a multiplexer to identify the start points of the individual frames in the incoming bit stream. The frame length is $1 + 8 \times 24 = 193$ bits, and the transmission rate—at points (b) and (c) of Fig. 1.5-3—is 193×8000 bits/second = 1544 kilobits/second (kb/s).

The El frame contains 32 eight-bit time slots, numbered TS_0 through TS_{31}—see Fig. l.5-4(*b*)—and serves $m = 30$ trunks. TS_1 through TS_{15} and TS_{17} through TS_{31} hold the octets for the samples of the trunks. Time slot TS_0 has a fixed eight-bit synchronization pattern, and time slot TS_{16} contains signaling information [4,8].

A El frame thus has $32 \times 8 = 256$ bits, and the bit rate on the transmission system is $8000 \times 256 = 2048$ kb/s.

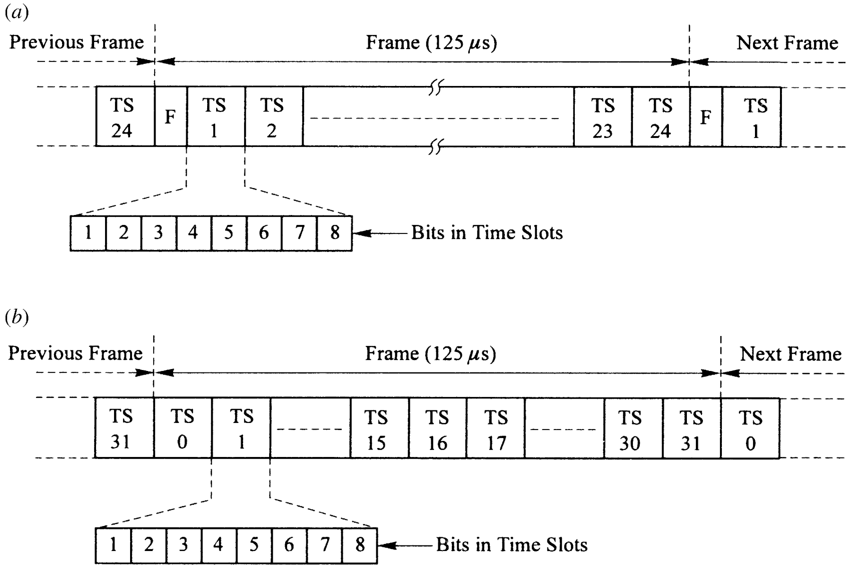


Figure 1.5-4. Formats of first-order PCM frames: (*a*) DS1 format ($m = 24$ channels) and (*b*) E1 format ($m = 30$ channels). TS, time slot.

Higher Order Multiplexes. As in FDM, higher order PCM multiplexes are formed by repeated multiplexing. The North American and the ETSI system have separate hierarchies [8].

1.5.3 Digital Transmission Systems [4,5]

First-order PCM multiplexes are carried on two copper wire-pairs in a conventional cable. The pulses attenuate and stretch as they traverse the cable and have to be regenerated by repeaters (R), located at intervals of 1 mile (Fig. 1.5-5). Since it is difficult to maintain transmission systems with large numbers of repeaters, the maximum length of these systems is about 200 miles.

The signals of higher order multiplexes can be carried on transmission systems of several types.

The North American T4M system carries a fourth-order PCM multiplex (4052 channels) on a pair of coaxial cables, with 1-mile repeater spacing. The T4M system is also limited to rather short trunks. Longer trunks are carried on microwave radio, or fiberoptic transmission systems.

In PCM microwave radio systems, the digital bit stream modulates a microwave carrier signal. *Phase-shift keying* (similar to DPSK modems—see Section 1.4.1) is a frequently used modulation form. Both terrestrial and satellite microwave transmission systems are in use. Like their analog counterparts, digital terrestrial radio links are divided into sections, with repeater stations at the section boundaries.

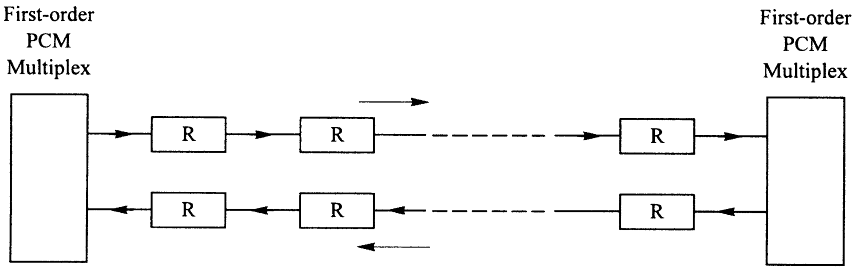


Figure 1.5-5. Transmission of first-order PCM multiplex on wire pairs. R, repeaters.

In optical fiber systems, the digital signal is in the form of lightwave pulses that propagate through a thin glass fiber (diameter on the order of 0.01 mm). At the sending end of the fiber, a laser diode converts the electrical pulses into "light" pulses. A photodiode at the receiving end converts the optical pulses back into electrical pulses. These transmission systems offer high-speed transmission with low attenuation and permit longer repeater spacing. Data rates of several gigabits/second are possible with repeater spacing of hundreds, even thousands, of kilometers, depending on the type of fiber.

Two standards for optical transmission exist:

1. Synchronous Optical Network (SONET)
2. Synchronous Digital Hierarchy (SDH)

Those standards specify a synchronous, framed transmission structure based on a hierarchical arrangement where lower order rates (called *tributaries*) are multiplexed into higher-order rates by byte interleaving, after adding overhead. SONET and SDH frames are transmitted at a rate of 8000 frames/second (every 125 μs), like T1 and E2 frames.

SONET is a North American standard, originally developed by Bellcore (now Telcordia), specified by ANSI [9,10], and fully compatible with T1 and multiples of T1 rates. SONET refers to an electrical signal as a *synchronous transport signal* (STS) and to its optical equivalent as an *optical carrier* (OC). Levels in the STS/OC hierarchy are called STS-n and OC-n with n = 1, 3, 12, 48, 192, and so on. The base rate is STS-1/OC-1 (51.84 Mbps) and can carry 28 DS-1 systems.

SDH, developed after SONET, is an international standard specified by ITU-T [11–13], which follows a similar hierarchical structure and is compatible with both T1 and E1 rates. Because most, if not all, SDH implementations carry only ETSI-defined data rates (E1 and multiples of E1 rates) as payload of the base rate, SDH is considered a European standard. SDH has been adopted by ETSI and is widely used outside North America. SDH signals are referred to as *synchronous transport modules* (STMs), and their hierarchical levels are called STM-n with n = 1, 4, 16, 64, and so on. The base rate is STM-l (155.52 Mbps) and can carry 63 E1 systems.

SONET is compatible with SDH starting with STM-3/OC-3, which corresponds to STM-1 (155.52 Mbps). STS-12/OC-12 then corresponds to STM-4, and so on.

1.5.4 Adaptive Differential Pulse Code Modulation

Since the introduction of PCM, several other coding techniques have been developed. The main objective of these developments is to lower the bit rate while maintaining the transmission quality. One of these methods, *adaptive differential pulse code modulation* (ADPCM), is being introduced in some telecommunication networks.

In ADPCM, four-bit digital samples, transmitted at 8000 samples/second, represent the *differences* in magnitude of two consecutive samples of the analog input signal. The coding is "adaptive": the relation between the magnitude of the analog quantity and the associated digital code is not fixed, but adapts automatically to the characteristics of the signal being encoded.

An ADPCM channel for a trunk requires $8000 \times 4 = 32$ kb/s and gives a speech quality almost equal to PCM. ADPCM doubles the channel (trunk) capacity of digital transmission systems. For example, by dividing the eight-bit octets of the E1 frame of Fig. 1.5-4(*b*) into two four-bit groups, a first-order E1 transmission system can serve 60 trunks.

1.6 SPECIAL TRANSMISSION EQUIPMENT

This section describes two types of special transmission equipment: *echo control* equipment and *circuit multiplication* equipment. Both of these exist in analog and digital versions.

1.6.1 Echoes

Figure 1.6-1 shows the transmission circuit (omitting the exchanges) for a typical long-distance connection with analog four-wire trunks. It consists of three parts. Parts 1 and 3 are two-wire circuits, containing the subscriber loops, and possibly two-wire trunks. Part 3 is a four-wire trunk. Hybrids H_1 and H_2 convert two-wire transmission into four-wire transmission.

When subscriber S_1 speaks, the speech signal leaves H_1 at port P, reaches port Q of H_2, and then travels on the two-wire circuit to listener S_2. However, a small part of the signal received at Q "leaks" to R, and thus returns to S_1, who hears an *echo* of his speech. The leakage occurs because hybrid circuits are "balancing" circuits. For leak-free operation, the impedance presented by the two-wire circuit (at port T of hybrid H_2) would have to match the design impedance of the hybrid for all voice-band frequencies. In each call, H_2 is connected to a different two-wire circuit, and impedances of these circuits (which vary with circuit length and physical cable characteristics) cannot be precisely controlled. Complete balance therefore never occurs in practice, and some echo is always present. When S_2 speaks, a similar echo is caused by reflections at H_1.

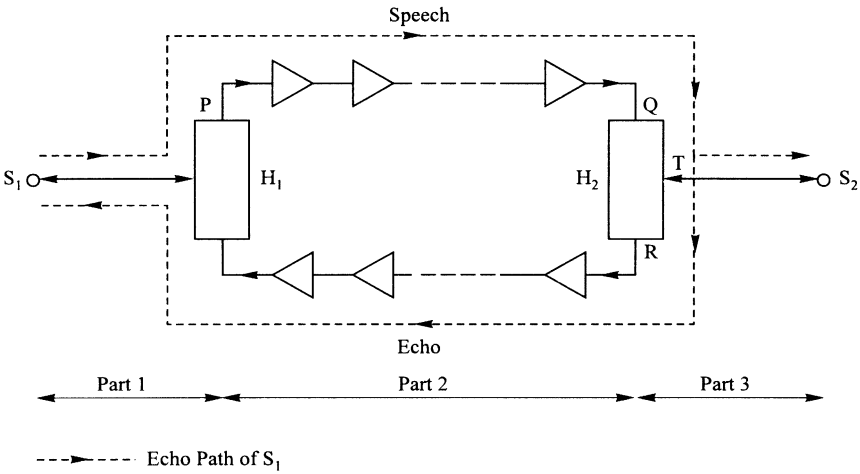


Figure 1.6-1. Long-distance connection.

The effect of echoes on a talker varies with the *echo delay* (the signal propagation time from talker S_1 to H_2 and back to S_1). Echoes delayed by less than 20 ms are barely noticeable, but echoes with larger delays rapidly become very disturbing to a talker. Echo delays increase with increasing length of the four-wire circuit, and trunks longer than 2000 miles (transcontinental, transoceanic, and satellite trunks) are equipped with echo control devices.

1.6.2 Echo Suppressors

Long four-wire analog trunks are equipped with echo suppressors [1,14]. Figure 1.6-2 shows echo suppressor units ES-A and ES-B, located at both ends of a trunk between exchanges A and B (the other exchanges in the connection are not shown).

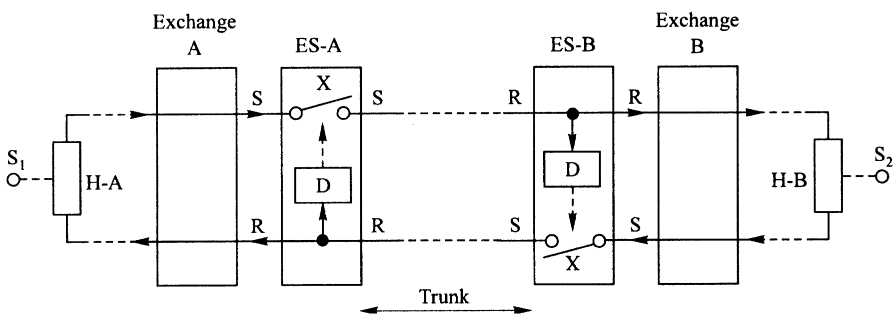


Figure 1.6-2. Echo suppressors on a long-distance analog trunk. D, Voiceband signal detector; X, switch; H, hybrid circuit; ES, echo suppressor.

Each suppressor unit has a detector (D) on its receive pair (R). When D detects the presence of a signal, the suppressor interrupts its send pair (S). For example, when S_1 speaks, ES-B interrupts its S path, and the leakage from hybrid H-B cannot return to S_1. ES-A protects S_2 from echoes in the same way.

When a suppressor has opened its send path, it "hangs over" (leaves the path open) for some 30–40 ms after the signal on the receive path has disappeared. This bridges the silent intervals between successive words of the distant speaker. Suppose now that S_1 has been speaking. The send path of ES-B is therefore open. When near-end subscriber S_2 starts to talk during a hangover, the initial part of the speech is clipped off. The effects of clipping become very noticeable on connections involving several echo-suppressed trunks in tandem. It is therefore desirable to avoid such connections. We shall see that some trunk signaling systems include indicators to make this possible.

Echo suppressor units are always installed in pairs—one at each end of a trunk. Some documents refer to the units as *half echo suppressors.*

1.6.3 Echo Cancelers

Echo cancelers [5,15], which are used on long PCM trunks, also provide protection against echoes but operate on a different principle. Figure 1.6-3 shows a pair of echo cancelers at the ends of a long-distance PCM trunk. The circuit between an echo canceler and its near-end subscriber is known as the "tail" and includes the subscriber line, hybrid circuit (H), an A/D and a D/A converter and can include other digital and analog trunks.

We assume that S_1 is speaking and explore the operation of canceler EC-B. The signals r, e, c, and s are PCM signals (sequences of octets). When S_1 speaks, a signal

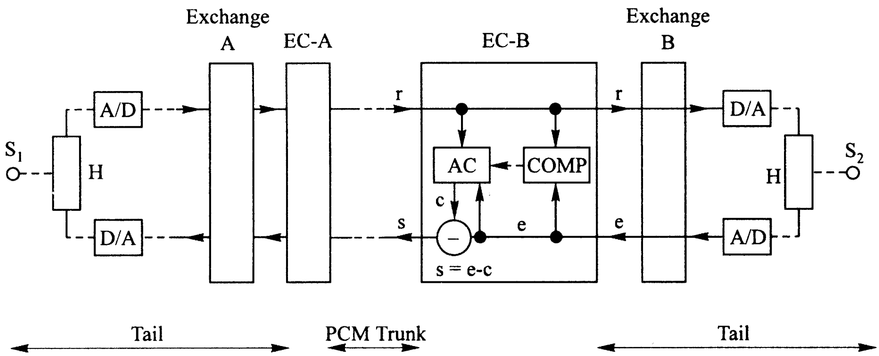


Figure 1.6-3. Echo cancelers on a PCM trunk. EC, Echo canceler; AC, adaptive circuit; COMP, comparator; A/D, analog-to-digital converter; D/A, digital-to-analog converter; H, hybrid circuit.

(r) is received on the receive pair of EC-B. During incoming speech, EC-B regards its tail as a two-port circuit that receives a signal (r) and returns an echo signal (e). Signal r also feeds the *adaptive circuit* (AC) in the canceler, which produces an output signal (c). Signal c is subtracted from signal e, and the results, s = e − c, is sent out on the send pair of EC-B.

The tail of an EC is different for each call. At the beginning of a call, the canceler automatically adjusts its adaptive circuit, AC, making the impulse transfer function of the circuit equal to that of the tail. This is done by comparing signals e and c. When the adaptation is complete, signals e and c are equal, and s = 0, which means that the echoes returning from the tail are being canceled. The adjustment takes place during the initial seconds of received far-end speech.

Comparator (COMP) compares the strengths of signals r and e. When r > e, the canceler is receiving speech from S_1 and can adjust circuit AC. However, when e > r, near-end subscriber S_2 is speaking, and no adjustments are made.

Echo cancelers have one significant advantage over echo suppressors: they do not clip the initial part of the near-end subscriber's speech. Therefore, more than one canceler-equipped trunk is allowed in a connection.

Digital Multiplexed Echo Cancelers. In the description above, the EC pair cancels echoes for one connection. Actual echo cancelers are digital devices that serve groups of *m* multiplexed PCM trunks. The canceler is connected to the output of a first-order PCM multiplex (Fig. 1.6-4) and processes the eight-bit PCM samples (octets) of the trunks, during their time slots in each PCM frame. The canceler has an adaptive circuit for each trunk, but its control circuitry is time-shared.

1.6.4 Circuit Multiplication

In a typical telephone conversation, each subscriber speaks during about 30% of the time, and both subscribers are silent during the remaining 40%. During the call, each channel of a four-wire trunk thus carries speech during only 30% of time. *Circuit multiplication equipment* allows a group of four-wire trunks to be carried on a four-wire transmission system with a smaller number of *bearer channels*, by making use of the "silent" intervals [5,14,16,17]. Circuit multiplication is economically

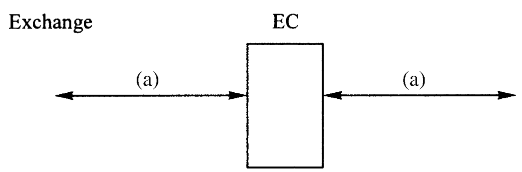


Figure 1.6-4. Echo cancelers for first-order digital time-division multiplexes (*m* channels). (a), Four-wire multiplexed circuits.

attractive in situations where transmission costs are high, for example, in transoceanic transmission systems.

Analog Circuit Multiplication Equipment. Circuit multiplication systems for analog four-wire trunks are generally known as *time assignment speech interpolation* (TASI) systems. Figure 1.6-5 shows two TASI units, at exchanges A and B, that concentrate m trunks to n (four-wire) bearer channels ($n < m$). In each TASI unit, the ($n \times m$) send switch allows up to n simultaneous unidirectional connections between the send circuits (S) of trunks and bearer channels. Likewise, the ($n \times m$) receive switch can connect up to n receive circuits (R) of bearer channels and trunks. Speech detectors are bridged across the S-pairs of the trunks. The controls in the TASI units communicate with each other on a pair of unidirectional control channels.

The assignments of bearer channels to trunks take place independently for each direction of transmission. The TASI operation for transmission from A to B is outlined below.

When a detector in TASI-A detects speech on the S-pair of a trunk, control- A seizes an available S-bearer channel and connects the trunk and the channel. It also informs control-B, identifying the trunk and the channel, and control-B then sets up a path between R-pair of the trunk and the R-bearer channel. This establishes a unidirectional path between the S-pair of the trunk at exchange A and the R-pair of the trunk at exchange B. At the end of a speech burst, an *overhang* timer, with an expiration time of about 400 ms, is started. If new speech energy is detected on the bearer channel while the timer is running, the timer is stopped. Expiration of the timer indicates a 400-ms silent interval on the channel. Control-A then releases the channel and informs control-B.

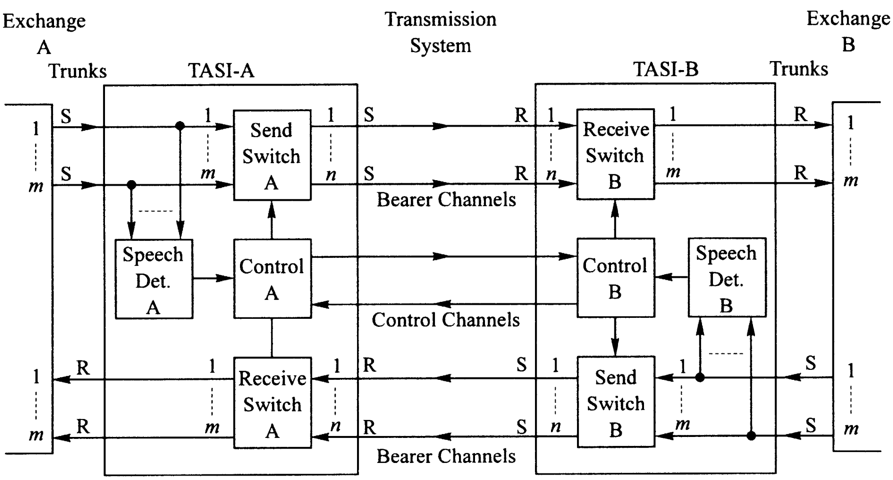


Figure 1.6-5. TASI equipment. R, Analog receive circuits; S, analog send circuits.

The setup of connections is very fast (typically, within 20 ms), and only a small part of the initial syllable of the first word in a sentence is "lost." This is barely noticeable to the listener.

Freeze-out. It can happen that all S-bearer chanels are occupied at the time that speech is detected on a previously silent trunk. In that case, the system has to wait until a bearer channel becomes available, and the initial part of a speech burst is "frozen out" (lost).

The ratio of bearer circuits to trunks is designed such that the probability of freeze-outs lasting more than 60 ms is below 1%. In TASI systems, typical concentration ratios m:n are on the order of 2:1.

TASI systems were designed during the years when speech was the only form of subscriber communication. Today, a small but increasing fraction of calls are used for facsimile and data transmission. During these calls, modem signals can be present continuously in one or both directions, thus occupying one or two bearer circuits on a full-time basis. To keep freeze-outs on the other trunks to an acceptably low level, it has become necessary to reduce the number of trunks that can be served by a TASI system.

Digital Circuit Multiplication Equipment. Digital circuit multiplication equipment operates on the same principle but has first-order multiplex interfaces with the exchange and the transmission system [17]. Moreover, most digital TASI systems convert the eight-bit PCM octets on the trunks to four-bit ADPCM groupings on the bearer channels. This doubles the channel capacity of the transmission system and reduces the per-channel transmission cost by another 50%. Figure 1.6-6 shows a typical configuration, where 150 PCM trunks (five first-order E1

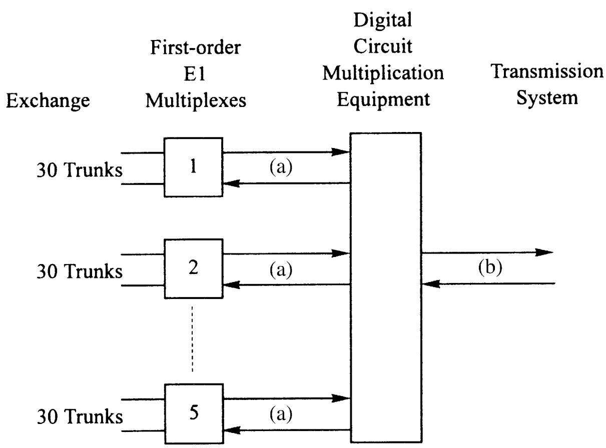


Figure 1.6-6. Digital circuit multiplication equipment. (a), 2048 kb/s E1 line (30 trunks); (b), 2048 kb/s E1 line (60 bearer channels).

multiplexes) are concentrated to 60 ADPCM bearer channels and carried on one E1 multiplex line.

1.7 EXCHANGES

1.7.1 Exchange Equipment

The primary function of an exchange is to establish and release temporary paths between two subscriber lines, between a subscriber line and a trunk, or between two trunks [4].

The two major equipment units in an exchange are shown in Fig. 1.7-1. The *switchblock* (or *switch fabric*) has a large number (up to some 100,000) of *ports* (P) to which subscriber lines and trunks are attached. The ports are interconnected by arrays of switches. By closing or opening appropriate sets of switches, paths across the switchblock can be set up or released. At any point in time, a multitude of these paths can be in existence. In Fig. 1.7-1, path 1 connects a line and a trunk, path 2 connects two lines, and path 3 connects two trunks.

The switchblock paths are set up and released on command from the *control equipment* (CE) of the exchange. CE sends commands to, and receives responses from, the switchblock via a *control channel* (CC) (shown as a dashed line).

1.7.2 Switchblocks

Analog Switchblocks. Until about 1975, all exchanges had analog switchblocks, implemented with electromechanical switching devices. In local exchanges, the switchblocks were two-wire (Fig. 1.7-2). The temporary paths in the switchblock, and the circuits attached to ports P, were two-wire bidirectional analog voiceband (300–3400 Hz) circuits (a).

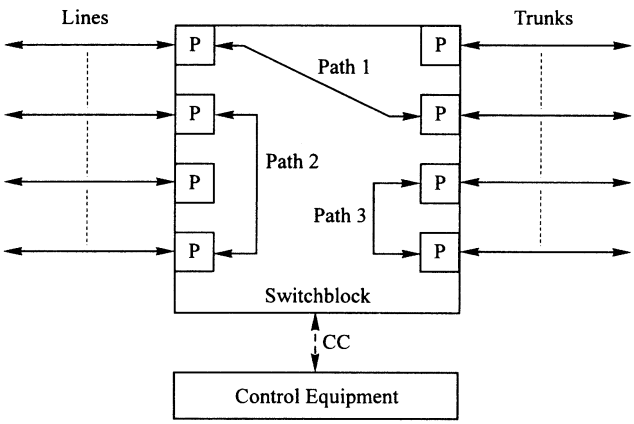


Figure 1.7-1. Exchange equipment. P, Port; CC, control channel.

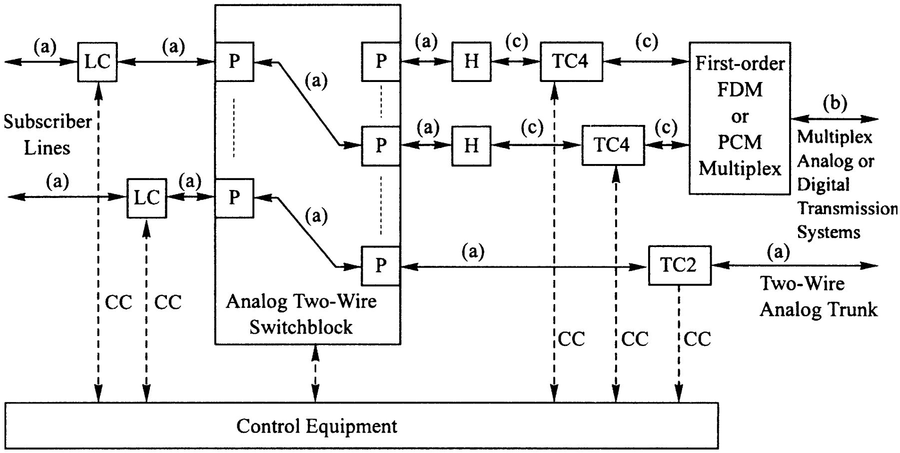


Figure 1.7-2. Local exchange with two-wire analog switchblock. P, Port; LC, line circuit; H, hybrid circuit; TC2, analog two-wire trunk circuit; TC4, analog four-wire trunk circuit; CC, control channel.

Subscriber lines and two-wire analog trunks (a) were attached to the ports of the switchblock via, respectively, two-wire *line circuits* (LCs) and two-wire trunk circuits (TC2). Multiplexed analog and PCM trunks (b) were converted, by first-order FDM and PCM multiplexes, to individual four-wire analog circuits (c). These circuits then passed through four-wire analog trunk circuits (TC4) and were converted by hybrids H into bidirectional analog two-wire circuits (a).

The switchblock and the line and trunk circuits were controlled by the control equipment of the exchange, via control channels (CCs). The line and trunk circuits play a role in the transmission and reception of signaling information (see Chapters 3 and 4).

The analog switchblocks in intermediate exchanges (tandem exchanges, toll exchanges, and international switching centers) provided four-wire analog paths between four-wire ports. That eliminated the need for hybrid circuits on multiplexed FDM and PCM trunks. Hybrid circuits are required on two-wire analog trunks, but the trunks attached to intermediate exchanges are predominantly four-wire trunks.

Digital Switchblocks. The introduction of digital time-division multiplex transmission systems has led to the development of exchanges with digital switchblocks. Most exchanges installed after 1980 have digital switchblocks that are implemented with integrated semiconductor circuits. These switchblocks are more compact, and more reliable, than their analog predecessors.

As shown in Fig. 1.7-3, the ports on digital switchblocks are usally four-wire first-order time-division digital multiplex ports (DMPs). The frame formats at points (b) are as shown in Fig. 1.5-4. The DMPs of exchanges in the United States have the DSI format (for $m = 24$ circuits). In most other countries, the DMPs have the E1 frame format (for $m = 30$ circuits).

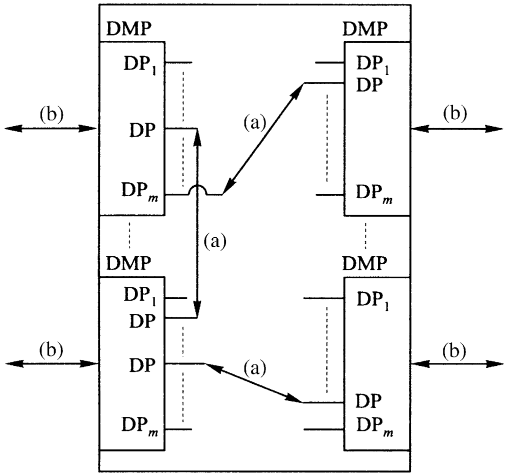


Figure 1.7-3. Paths in digital switchblock. (a), 64 kb/s four-wire paths; (b), first-order digital multiplex circuits (*m* channels); DMP, digital multiplex port.

The paths in the switchblock are 64 kb/s, four-wire circuits (a), The *m* multiplexed circuits entering a digital multiplex port DMP are segregated into *m* individual 64-kb/s digital ports (DPs), and the switchblock paths connect pairs of these ports.

Attachment of Lines and Trunks. (See Fig. 1.7-4.) T1 or E1 transmission systems for digital trunks (b) can be attached directly to the DMPs. Each DMP has a control channel (CC), on which the control equipment of the exchange sends and receives signaling information for the *m* multiplexed channels.

Subscriber lines arrive at the exchange as two-wire analog circuits (c). They pass through two-wire line circuits (LCs) and hybrid circuits (H) and enter first-order PCM multiplexers (Section 1.5.2). The outputs of the multiplexers are four-wire digital multiplexed circuits (b) that serve *m* lines and are connected to DMPs.

Each line circuit also has a control channel (CC) to the control equipment of the exchange.

Analog trunks arriving on FDM transmission systems are first demultiplexed by FDM multiplexes (Section 1.4.5) into individual four-wire analog circuits (d). These circuits then pass four-wire analog trunk circuits (TC4) and enter PCM multiplexes, where they are converted to a first-order digital multiplex circuit (b).

1.7.3 Control Equipment

The control equipment [1,4] of early exchanges was implemented with electromechanical devices. Around 1955, it became clear that digital computers could be used to control exchanges. The speed and sophistication of computers at that time already greatly exceeded the capabilities of electromechanical exchange control

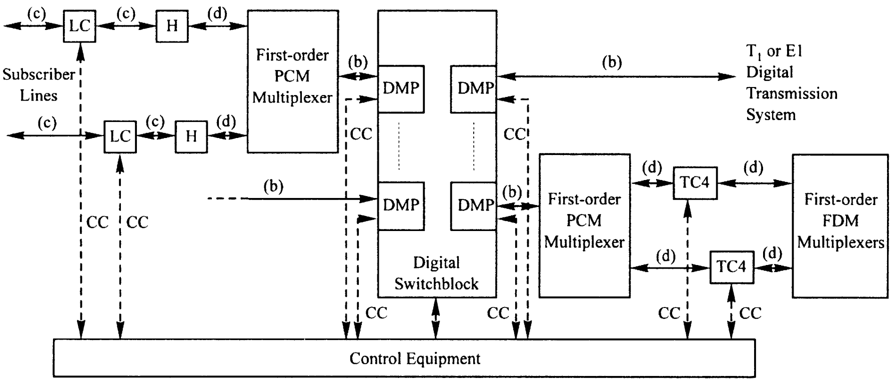


Figure 1.7-4. Exchange with digital switchblock. LC, Line circuit; H, hybrid circuit; DMP, digital multiplex port; TC4, analog four-wire trunk circuit; CC, control channel.

equipment. The first *stored-program controlled* (SPC) exchange was placed in commercial operation in 1965 [1].

The main elements of a stored-program exchange control system are shown in Fig. 1.7-5. A processor (or a group of processors) performs all exchange actions, controlling the switchblock, the line and trunk circuits, and other exchange equipment, by executing instructions in its program. The control channels for these equipment units are connected to a bus on the processor.

Processor memory is divided into semipermanent memory and temporary memory. The semipermanent memory stores the programs for call processing, charging, and exchange maintenance procedures, and tables with data for digit analysis, call routing and charging, and so on.

Temporary memory stores data that are changed frequently: for instance, information on the status of the subscriber lines and trunks, or records of currently existing calls.

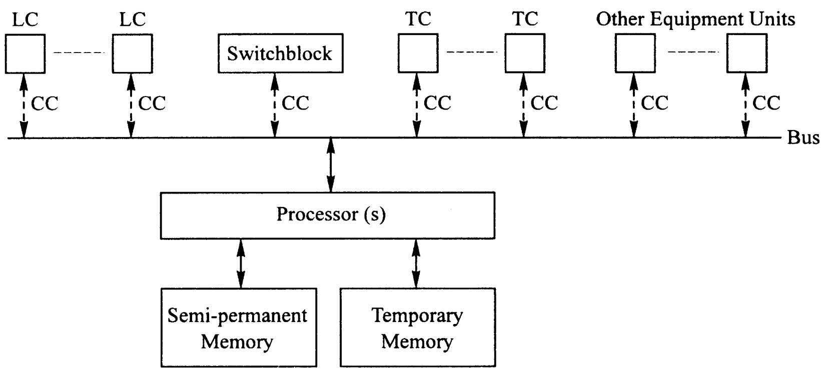


Figure 1.7-5. Stored-program exchange control. LC, Line circuits; TC, trunk circuits; CC, control channels.

The increased logical power of SPC exchanges has allowed the inclusion of features that were not available in earlier exchanges. These fall into two main categories:

- New services for subscribers (call forwarding, call waiting, three-way calling, etc.).
- Procedures to facilitate the administration, operation, and maintenance of the exchange by the telecom. These include routine checks on the exchange equipment, periodic testing of attached lines and trunks, the generation of data on the traffic handled by the exchange and its trunk groups, and the production of equipment trouble reports.

1.8 ACCESS NETWORKS AND LINE CONCENTRATORS

An *access network* (AN) is the part of the telephone network that connects the subscriber premises to the *local exchange* (LE). Initially, access networks consisted of subscriber lines (copper wire pairs) that terminated at the individual subscriber premises on one side and at the local exchange on the other side. Because of the cost involved in deploying the massive volume of copper wires needed to connect all local users to the LE, *remote line concentrators* (RLCs), sometimes known as *pair gain systems* or *loop carriers*, are often interposed between the subscriber premises and the LE. Concentrators, located close to the subscriber premises, collect a large number of line loops and reduce the volume of wires that go all the way from subscribers to the LE.

Various techniques are used for concentration, all taking advantage of the fact that not all subscribers are involved in a call at the same time. The earliest types of RLCs were simple devices that served N subscriber lines with a pool of M wire pairs ($M < N$) between the concentrator and the LE, assigning an idle pair from the pool to a subscriber for the duration of a call and then returning the pair to the pool at the end of the call (Fig. 1.8-1).

The telephone traffic for a residential line during the busy hour has been traditionally around 10% or 0.1 erlangs (although dial-up Internet service has tended to raise average holding times), and the concentration ratio N:M is typically around 4:1.

The more recent line concentrators are called *digital line concentrators* or *digital loop carriers* (DLCs), which multiplex the calls of N subscribers onto a digital multichannel link, such as a T1 ($M = 24$) or E1 ($M = 30$) line connecting the concentrator to the LE (Fig. 1.8-1(*a*)). Each call is assigned to a channel (time slot—TS). With this technique, two types of multiplexing are possible:

- Fixed assignment of time slots to user interfaces
- Dynamic (statistical) assignment of time slots to user interfaces

With fixed assignment, the number of channels on the digital link is equal to or greater than the number of line interfaces ($M \geq N$), and there is no blocking since each user interface is permanently connected to a channel. The "pair gain" is due

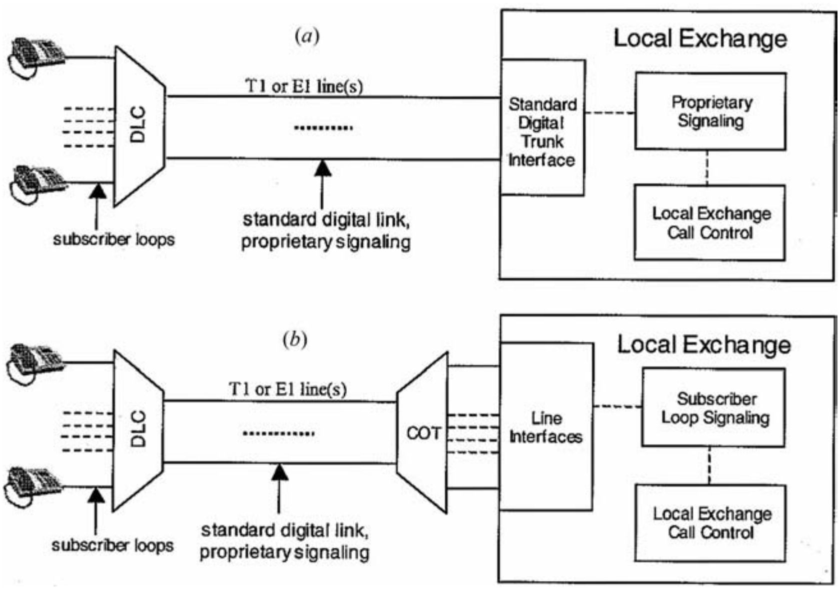


Figure 1.8-1. Digital line concentrators.

to the fact that one physical link (the T1 or E1 line) is carrying the traffic of 24 or 30 concurrent connections. With statistical assignment, fewer TSs are available than the number of line interfaces ($M < N$) and further savings in facilities are achieved through concentration. The drawback is that blocking is possible in the case of unusually high traffic. Multiple E1/T1 lines may be used on one RLC/DLC interface for reliability or if the traffic warrants it.

At the local exchange, the E1/Tl lines can be connected directly to a digital trunk interface, or to a demultiplexer that recreates the subscriber line interfaces as they would appear if the loops had gone all the way from the subscriber premises to the LE. The demultiplexer is known in North America as a *central office terminal* (COT—Fig. 1.8-1(*b*)).

Signaling is required between the concentrator and the LE or COT to handle DLC functions. In the past, each equipment supplier provided its own proprietary signaling protocols, and the COT was the only way to ensure that a DLC would interwork with any type of local exchange, regardless of the supplier. The situation started to change in the 1980s with standardization efforts to "open" the interface between DLC and LE and to eliminate the need for the COT. The new types of DLCs are known as *access systems*, a term that is associated with a standard digital trunk interface and a standard signaling protocol (Fig. 1.8-2). Signaling for access systems is described in Chapter 6.

Access systems have led to a local exchange configuration where the LE retains the call control function for lines but has no physical line interfaces. All physical interfaces are standard TDM digital trunks and the physical interface to local loops is handled by

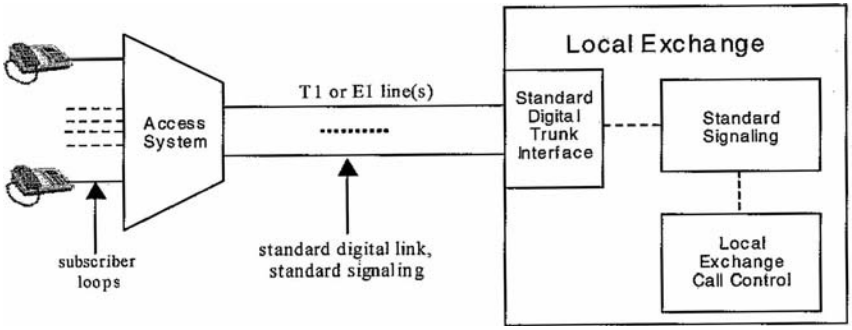


Figure 1.8-2. Access systems.

access systems. Functionally, access systems can be viewed as open-interface line units, which can be deployed remotely or collocated with the LE (when loop lengths are very short), and which can be procured independently of the LE.

1.9 ACRONYMS

A/D	Analog-to-digital
AC	Area code, adaptive circuit
ADPCM	Adaptive differential pulse code modulation
ALG	Access line group
AN	Access network
CC	Country code, control channel
CE	Control equipment
CIC	Carrier identification code
COT	Central office terminal
D/A	Digital-to-analog
DLC	Digital line concentrator, digital loop carrier
DMP	Digital multiplexer port
DPSK	Differential phase-shift keying
DS1	North American first-order pulse code modulation multiplex
EC	Exchange code, echo canceler
ES	Echo suppressor
ETSI	European Telecommunications Standards Institute
FDM	Frequency-division multiplexing
FSK	Frequency-shift keying
H	Hybrid circuit
IC	Interexchange carrier, identification code
IXC	Interexchange carrier
IN	International number
ISC	International switching center

ISDN	Integrated Services Digital Network
ITU	International Telecommunication Union
ITU-T	Telecommunication Standardization Sector of ITU
LATA	Local access and transport area
LC	Line circuit
LE	Line equipment, local exchange
LEC	Local exchange carrier
LN	Line number
MP	Multiplex port
MSC	Mobile switching center
NANP	North American numbering plan
NN	National number
NPA	Numbering plan area
OC	Optical carrier
P	Single port
PBX	Private branch exchange
PCM	Pulse code modulation
PSTN	Public switched telecommunications network
R	receive circuit of four-wire circuit
RF	Radiofrequency
RLC	Remote line concentrator
S	Subscriber, send circuit of four-wire circuit
SCN	Switched-circuit network
SDH	Synchronous Digital Hierarchy
SL	Subscriber line
SN	Subscriber number
SONET	Synchronous Optical Network
SPC	Stored-program controlled
STM	Synchronous transport module
STS	Synchronous transport signal
T	Trunk
TASI	Time assignment speech interpolation
TC	Analog trunk circuit
TC2	Analog two-wire trunk circuit
TC4	Analog four-wire trunk circuit
TDM	Time-division multiplexing
TG	Trunk group
VPN	Virtual private network

1.10 REFERENCES

1. *Engineering and Operations in the Bell System*, 2nd ed., AT&T Bell Laboratories, 1984.
2. *The International Public Telecommunication Numbering Plan*, Rec. E.164, ITU-T, Geneva, 1997.

3. G. R. Ash, Design and Control of Networks with Dynamic Nonhierarchical Routing, *IEEE Commun*., **28**, Oct. 1990.
4. R. L. Freeman, *Telecommunication System Engineering*, 2nd ed., Wiley-Interscience, Hoboken, NJ, 1989.
5. *Transmission Systems for Communications*, 5th ed., Bell Telephone Laboratories, 1982.
6. *Data Communication Over the Telephone Network*, Rec. V.10–V.32, ITU-T, Geneva, 1988–2000.
7. G. H. Bennett, *PCM and Digital Transmission*, Marconi Instruments Ltd., UK, 1978.
8. *Digital Terminal Equipments*, Rec. G.700–G.798, ITU-T, Geneva, 1998–2003.
9. *Synchronous Optical Network (SONET)—Data Communication Channel Protocol and Architectures*, ANSI T1.105.04-1995 (R2001), New York, July 2001.
10. *Synchronous Optical Network (SONET)—Sub-STS-1 Interface Rates and Formats Specification*, ANSI Tl.105.07-1996 (R2001), New York, July 2001.
11. *Network Node Interface for the Synchronous Digital Hierarchy (SDH)*, Rec. G.707, ITU-T, Geneva, Oct. 2000.
12. *Sub STM-0 Network Node Interface for the Synchronous Digital Hierarchy (SDH)*, Rec. G.708, ITU-T, Geneva, July 1999.
13. *Interfaces for the Optical Transport Network (OTN)*, Rec. G.709, ITU-T, Geneva, Mar. 2003.
14. J. M. Frazer, D. B. Bullock, and N. G. Long, Overall Characteristics of a TASI System, *Bell System Tech. J*., **51**, July 1962.
15. *International Telephone Connections and Circuits*, Rec. G.101–G.181, ITU-T, Geneva, 1998–2003.
16. D. A. Bardouleau, Time Assignment Speech Interpolation Systems, *British Telecom. Eng*., **4**, Oct. 1985.
17. M. Onufry, The New Challenge of DCME, *Telecomm. J*., **55**, Dec. 1998.

2

INTRODUCTION TO SIGNALING

2.1 OVERVIEW

This section presents a brief historical outline of signaling. The earliest telephone exchanges were "manual" switchboards, in which all calls were set up and taken down by operators. Signaling between subscribers and operators was limited to *ringing*. To make a call, the subscriber would send a ringing signal. This alerted an operator, who would connect her telephone to the calling line and ask for the called number. The operator then would connect her telephone to the called line and ring the line. After answer by the called party, the operator would establish the connection.

Signaling as we know it today started around 1890, with the invention, by Almon B. Strowger (a Kansas City undertaker), of an automatic switchboard that could receive the called number dialed by the calling subscriber and would then automatically set up the connection. During the past 100 years, signaling applications and technology have evolved in parallel with the developments in telecommunications.

2.1.1 Early Signaling

Signaling in the period from 1890 to 1976 had three main characteristics. In the first place, its application was limited to *plain old telephone service* (POTS): the setup and release of connections between two subscribers. In the second place, the signals were carried by the same circuit (subscriber line, trunk) that carried the speech during the call. This type of signaling is known as *channel-associated signaling* (CAS) [1,2]. Finally, signaling took place only between a subscriber and the local exchange (*subscriber signaling*), and between the exchanges at the two ends of a trunk (*interexchange signaling*).

Initially, automatic telephony was possible only for calls between subscribers served by the same exchange, which required subscriber signaling only. Later on,

Signaling in Telecommunication Networks, Second Edition, by John G. van Bosse and Fabrizio U. Devetak

it became possible to dial calls between subscribers served by nearby exchanges. These calls also required interexchange signaling. National long-distance calls needed operator assistance until the 1950s, when *direct distance dialing* (DDD) was introduced. *International direct distance dialing* (IDDD), which requires signaling on international trunks, became possible in the 1960s [3].

Channel-associated call-control signaling is still widely used today. However, beginning in 1976, other forms and applications of signaling have made their appearance in telecommunication networks. They are briefly outlined below.

2.1.2 Common-Channel Signaling

Common-channel signaling (CCS), introduced in 1976, was developed as an alternative form of call-control signaling for trunks [2,4,5]. In CCS, signaling information is not carried by the individual trunks. Instead, a *signaling network*, consisting of *signaling data links* (SDLs) and *signal transfer points* (STPs), transfers digital *signaling messages* between the exchanges. In Fig. 2.1-1, trunk groups TG_1, TG_2, and TG_3 have common-channel signaling, and the signaling network consists of one signal transfer point (STP) and the signaling data links SDL_A, SDL_B, and SDL_C. A call-control message from exchange A to exchange B for a trunk in TG_1 traverses SDL_A, the signal transfer point, and SDL_B. In this example, each signaling link carries messages for all CCS trunks that are attached to an exchange. For example, SDL_A carries messages for the trunks in groups TG_1 and TG_2. We say that SDL_A is the "common" signaling channel for the trunks in these groups.

Common-channel signaling became possible after the introduction of stored-program controlled exchanges. We shall see that CCS is more powerful, flexible, and also faster than CAS. CCS was introduced rapidly in national telecommunication

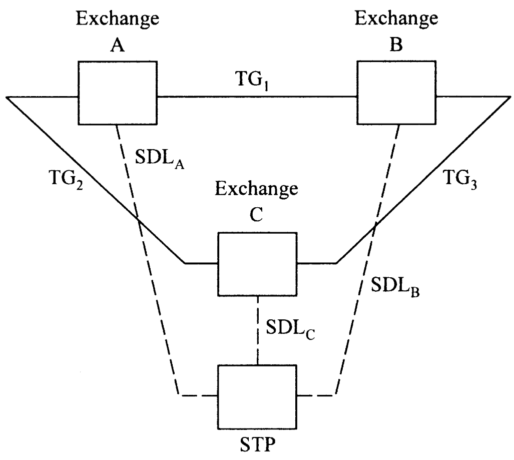


Figure 2.1-1. Common-channel call-control signaling. $TG_{1,2,3}$, Trunk groups; $SDL_{A,B,C}$, signaling data links; STP, signal transfer point.

networks and in the international network. However, the replacement of CAS is not yet complete, and CAS and CCS coexist in many networks.

2.1.3 Other Applications of Common Channel Signaling

Since 1980, CCS is also being used for other applications. In Fig. 2.1-2, a *service control point* (SCP) and an *operations, administration, and maintenance* (OAM) *center* also have signaling data links, and exchanges can send messages to, and receive messages from, these network entities. Procedures that involve signaling between an exchange and a SCP, or between an exchange and an OAM center, are known as *transactions*.

The SCPs in a network support *intelligent network* (IN) services [5–7]. These services require information that cannot be stored conveniently in exchanges. A well-known IN service is 800 calling. The 800 numbers are not in the conventional format of national numbers and cannot be used by the exchanges to route calls to their destinations. Therefore, when an exchange receives a call with an 800 number, it starts a transaction with the SCP, requesting the translation of the received 800 number. The SCP then translates the number into the national number of the called party and returns that number to the exchange, which then proceeds to set up the connection.

The OAM centers allow centralized operation, maintenance, and administration of the network. For example, there are transactions in which the OAM center requests an exchange to test a particular trunk and to report the test results. Other

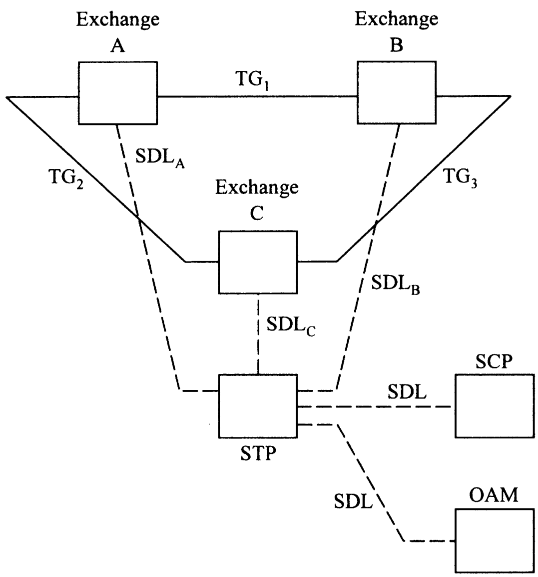


Figure 2.1-2. Common-channel signaling for transactions involving an exchange and a service control point (SCP) or an operations, administration, and maintenance (OAM) center.

transactions enable the OAM center to verify and change the subscriber and routing data that are stored in the exchanges.

2.1.4 Signaling in Cellular Mobile Networks

Cellular mobile telecommunications were introduced in the 1980s and are widely used today [8]. Figure 2.1-3 shows a cellular mobile network. Exchanges in mobile networks are known as *mobile switching centers* (MSCs). The MSC has one or more trunk groups (TGs) to exchanges in the *public switched telecommunication network* (PSTN). The service area of a MSC is divided into a number of cells. Each cell has a *base station* (BS) that contains microwave radio equipment and has trunks and data links to the MSC.

A *mobile station* (MS) in a cell communicates with the base station of that cell on a radiofrequency channel, of which there are two kinds. A *voice channel* (VC) of a base station is permanently connected to a trunk between the BS and the MSC. A *control channel* (CC) of a base station is permanently connected to a *data link* between the BS and the MSC.

In Fig. 2.1-3, mobile MS_A is engaged in a call. The MSC has allocated trunk T_x, and the associated voice channel VC_x, to the call. T_x and VC_x transfer the speech, and the signaling, between the mobile and the MSC.

The signaling for all mobiles in a cell that are not involved in a call is carried on a control channel of the cell and its associated data link. For example, mobile MS_B is

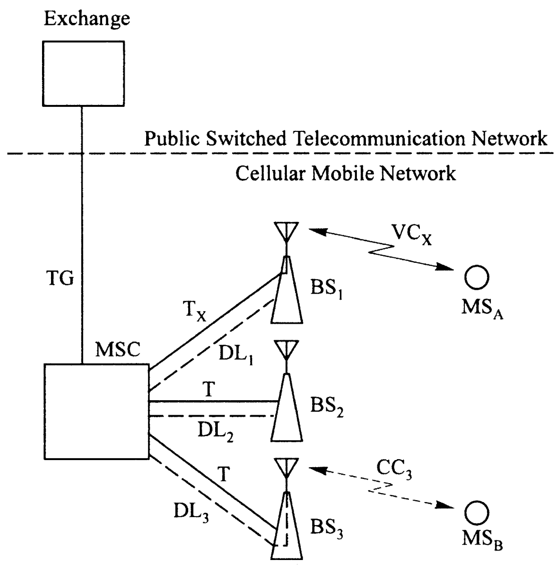


Figure 2.1-3. Signaling in cellular mobile networks. MSC, Mobile switching center; MS, mobile station; VC, voice channel; CC, control channel; T, trunk; DL, data link; BS, base station.

not involved in a call, and the signaling between the mobile and the MSC takes place on DL_3 and CC_3. The signaling between the mobiles and the MSC is thus a combination of channel-associated and common-channel signaling.

In addition, signaling in cellular mobile networks includes transactions between mobile switching centers and mobile network databases. One group of these transactions allows a mobile to obtain service while it is *roaming* outside its "home" cellular system.

2.1.5 Digital Subscriber Signaling

The public switched telecommunication networks in some countries have been converted (often only partially) into *Integrated Services Digital Networks* (ISDNs) [2,9,10]. An ISDN serves conventional (analog) subscribers and ISDN *users*. Digital ISDN users can communicate with each other in two modes. In circuit mode, the network sets up a dedicated connection for the call, which can be used for voice and data communication. In packet-mode communication (Chapter 20), the users communicate with short bursts of data, called *packets*.

There are 64-kb/s ISDN terminals (terminal equipment—TE) of several types, for example, digital telephones, high-speed facsimile terminals, and high-speed computer modems (Fig. 2.1-4). A *digital subscriber line* (DSL) connects the user's TE to the local exchange. ISDN DSLs are two-wire or four-wire circuits that allow simultaneous information transmission at 144 kb/s in both directions, although, due to *overhead* bits, the total transmission rate is somewhat higher. Signaling on ISDN lines is message oriented and is described in Chapter 10.

A more recent development is *asymmetric digital subscriber lines* (ADSLs) and less common variations thereof, such as high bit-rate DSL (HDSL) and symmetric DSL (SDSL), collectively referred to as xDSL. xDSL technology uses high-speed digital modems on regular subscriber loops to create data channels that can achieve rates of several megabits per second, depending on the length of the loop. An ADSL modem creates two bidirectional data channels, one a regular (narrowband) voice channel and the other a high-speed packet data channel. The ADSL data channel has a much higher bit rate in the direction from the local exchange

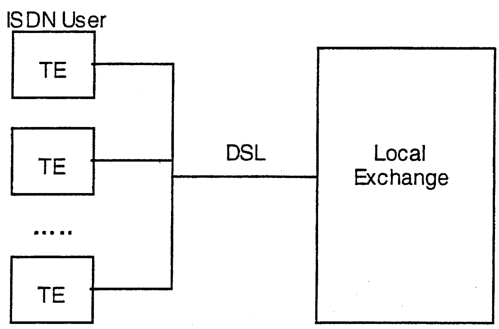


Figure 2.1-4. ISDN digital subscriber line.

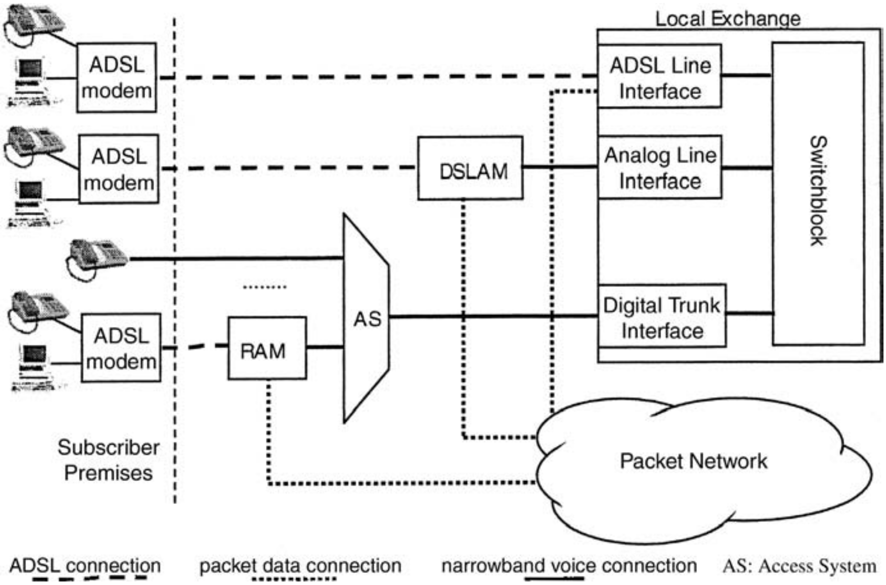


Figure 2.1-5. ADSL configurations.

to the subscriber than from the subscriber to the local exchange, hence the term "asymmetric," reflecting the fact that in normal communication a far larger volume of data is downloaded from the network than uploaded to it.

The two bidirectional channels are split at the local exchange, where the voice channel is connected to the switchblock and handled like any other subscriber line interface, and the data channel is connected to the packet data network (typically ATM—Chapter 22). A common arrangement is to terminate the ADSL at a *digital subscriber line access multiplexer* (DSLAM) collocated with the local exchange, but solutions where the local exchange has integrated ADSL interfaces in its line units are also available. With access systems (Section 1.8 and Chapter 6), which cannot transport an ADSL, a *remote access multiplexer* (RAM) may be interposed at the remote location, to split the voice channel (which connects to the AS) from the data channel (which connects to the packet network). ADSL configurations are shown in Fig. 2.1-5.

A similar approach to xDSL is being used by cable companies, who use the existing coaxial cable access for TV service to offer voice and high-speed data services to subscribers (cable modems).

The general trend is for xDSL and cable modems to replace ISDN as a high-speed digital subscriber interface.

2.1.6 Multiple Interexchange Signaling Systems

For economic reasons, telephone exchanges are kept in service for about 20 years. That means that both the national and the international networks contain different generations of exchanges. Some old-technology exchanges can handle only older interexchange signaling systems and cannot be upgraded to accommodate newer forms of signaling. Modern exchanges using SPC technology (Section 1.7) by and large can be retrofitted to handle newer signaling systems but, depending on their vintage, the process can require a nontrivial investment and long lead times. In some countries, such as the United States, telecoms have shifted a significant portion of the capital investment previously earmarked for new exchanges to equipment based on packet technology (Chapter 20), which requires a whole new set of signaling systems.

As a consequence, a connection routed via one or more intermediate exchanges of different vintages may involve trunks with different interexchange signaling systems. Call-control at such exchanges has to include procedures for signaling interworking between the different signaling systems. One aspect of interworking is the conversion of the formats of individual signals and/or messages. More difficult problems arise when a signal, or an information element in a common-channel signaling message, exists in a new system but not in an older system. Signaling interworking functions can be quite complex and have to be designed with care.

2.2 STANDARDS FOR SIGNALING SYSTEMS

The equipment in a telecommunication network is usually purchased from several manufacturers. In order to ensure that equipment from different suppliers can be interconnected without problems, the telecoms have developed (and continue to develop) standards that are documented in *specifications* and that have to be met by the manufacturer's equipment. Specifications cover signaling and many other aspects of telecommunications. The information in this book is based largely on signaling standards published by the organizations mentioned below.

2.2.1 North American Organizations

Alliance for Telecommunications Industry Solutions (ATIS). ATIS, originally called the Exchange Carrier Standards Association (ECSA), was created after divestiture to ensure that the North American network would continue to operate as an integrated entity [11]. With a membership that includes all the major telecoms and equipment suppliers, ATIS manages requirements through an entity called the *T1 Committee*, responsible for generating drafts of telecommunication standards. Proposed standards are submitted to the American National Standards Institute (ANSI) and published as American National Standards for Telecommunications.

Telcordia. Telcordia Technologies is an independent company that had its origins as the requirements coordination arm of the Bell System, when it was called Bellcore. Prior to divestiture in 1984, the Bell System was the dominant telecom and equipment manufacturer and set the de facto standards for the North American telecommunication network [12].

Bellcore became an independent company in 1984 and was renamed Telcordia Technologies in 1997. The company is a source of specifications for interfaces, capabilities, and performance in the form of Generic Requirement (GR) documents.

TIA/EIA. Standards for cellular mobile systems in the United States are established by subcommittees of the TR.45 Committee. They are published jointly by the Telecommunications Industry Association (TIA) and the Electronic Industries Alliance (EIA), as TIA/EIA standards.

2.2.2 ITU-T

Historically, the telecoms that operated individual national networks established standards independently of each other. With the advent of international telecommunications, the need for international standards arose and led to the establishment of international standards organizations.

The CCITT (International Telegraph and Telephone Consultative Committee) was established in 1956 as a part of the International Telecommunication Union (ITU), headquartered in Geneva, Switzerland [13]. As a result of a major reorganization of ITU, in 1993 the CCITT became the ITU-T, the Telecommunication Standardization Sector of ITU.

ITU-T Study Groups review draft standards and, when approved, standards become *recommendations*. Signaling standards are the responsibility of Study Group 11.

CCITT standards used to be published as a set of books identified by a color, the last set of which, the *Blue Book*, was published in 1989. ITU-T standards are now published individually on paper or in computer file form.

While ITU-T recommendations are intended primarily for the international network, many national telecoms have adopted them—sometimes with country-specific modifications—for use in their respective networks.

2.2.3 ETSI

European telecoms established the Conference of European Postal and Telecommunications Administrations (CEPT) in 1959 [14,15]. The standardization activities were transferred in 1998 to a new organization, the European Telecommunications Standards Institute (ETSI). The work is carried out by a number of technical committees, one of which deals with signaling protocols and switching (SDS). ETSI standards are published as *European Telecommunications Standards* (ETS).

2.2.4 Other National Organizations

Other organizations that handle telecommunication requirements at a national level are:

- Telecommunication Technology Committee (TTC), in Japan
- Association of Radio Industries and Businesses (ARIB), in Japan

- Communications Standards Association (CCSA), in China
- Telecommunication Technology Association (TTA), in Korea

2.2.5 Other International Organizations

Other international organizations have been created to coordinate and sponsor requirements efforts for newer and rapidly evolving networks.

Third Generation Partnership Project (3GPP). 3GPP was started in 1998, at the instigation of ETSI, to coordinate requirements for third generation wireless communication evolving from GSM (Chapter 12). 3GPP members, as of 2005, were ARIB, CCSA, ETSI, ATIS, TTA, and TTC.

Third Generation Partnership Project 2 (3GPP2). 3GPP2 started in 1998, in parallel with 3GPP, to coordinate requirements for third generation wireless communication evolving from IS-95 (Chapter 13). 3GPP2 members, as of 2005, were ARIB, CCSA, TIA, TTA, and TTC.

Internet Engineering Task Force (IETF). The IETF is the protocol standardization organization for the Internet (Chapter 20). It was established in 1986 under the auspices of the Internet Society (ISOC), a professional, nongovernmental organization that, in its own words, is "an international group dedicated to making the Internet work seamlessly around the world." IETF standards are called RFCs (Request For Comments) and are posted on the IETF website.

MFA Forum/ATM Forum. The MFA Forum is an international organization that resulted from the merger of three, originally separate, entities: the MPLS Forum, the Frame Relay Forum, and the ATM Forum. The MPLS Forum merged with the Frame Relay Forum in 2003, and with the ATM Forum in 2005. The MFA Forum coordinates requirements, focusing on implementation issues, for ATM (Chapter 22), as well as other packet technologies such as MPLS and Frame Relay (not covered in this book).

2.3 ACRONYMS

3GPP	Third Generation Partnership Project
3GPP2	Third Generation Partnership Project 2
ADSL	Asymmetric digital subscriber line
ANSI	American National Standards Institute
ARIB	Association of Radio Industries and Businesses
AS	Access system
ATIS	Alliance for Telecommunications Industry Solutions
ATM	Asynchronous transfer mode
BS	Base station

CAS	Channel-associated signaling
CC	Control channel
CCS	Common-channel signaling
CCSA	Communications Standards Association (China)
DDD	Direct distance dialing
DSL	Digital subscriber line
DSLAM	Digital subscriber line access multiplexer
EIA	Electronics Industries Alliance
ETSI	European Telecommunications Standards Institute
IDDD	International direct distance dialing
IETF	Internet Engineering Task Force
IN	Intelligent network
ISDN	Integrated Services Digital Network
ITU	International Telecommunication Union
ITU-T	Telecommunication Standardization Sector of ITU
MFA	MPLS, Frame Relay, ATM
MPLS	MultiProtocol Label Switching
MS	Mobile station
MSC	Mobile switching center
OAM	Operations, administration, and maintenance
POTS	Plain old telephone service
PSTN	Public switched telephone network
RAM	Remote access multiplexer
SCP	Service Control Point
SDL	Signaling data link
STP	Signal transfer point
TE	Terminal equipment
TIA	Telecommunications Industry Association
TTA	Telecommunication Technology Association (Korea)
TTC	Telecommunication Technology Committee (Japan)
VC	Voice channel

2.4 REFERENCES

1. S. Welch, *Signalling in Telecommunications Networks*, Peter Peregrinus Ltd., Stevenage, UK, 1981.
2. R. L. Freeman, *Telecommunication System Engineering*, 2nd ed., Wiley–Interscience, Hoboken, NJ, 1989.
3. *A History of Engineering & Science in the Bell System*, Bell Telephone Laboratories, 1982.
4. Special Issue on CCIS, *Bell Syst. Tech. J.*, **57**, No. 2, 1978.
5. R. Modaressi and R. A. Skoog, Signaling System No. 7: A Tutorial, *IEEE Commun. Mag.*, **28**, No. 7, July 1990.
6. S. M. Boyles, R. L. Corn, and L. M. Mosely, Common Channel Signaling: The Nexus of an Advanced Communications Network, *IEEE Commun. Mag.*, **28**, No. 7, July 1990.

7. R. B. Robrock, The Intelligent Network Evolution in the United States, *Conference Proceedings, Singapore ICCS'90*, Elsevier, Amsterdam, 1990.
8. W. C. Y. Lee, *Mobile Cellular Telecommunications*, McGraw-Hill, New York, 1995.
9. P. K. Verma, ed., *ISDN Systems*, Prentice Hall, Englewood Cliffs, NJ, 1990.
10. T. B. Bell, Telecommunications, *IEEE Spectrum*, **29**, No. 2, 1992.
11. A. K. Reilly, A U.S. Perspective on Standards Development, *IEEE Commun. Mag.*, **32**, No. 4, Jan. 1994.
12. *Notes on the Network*, AT&T, New York, 1980.
13. T. Irmer, Shaping Future Telecommunications: The Challenge of Global Standardization, *IEEE Commun. Mag.*, **32**, No. 4, Jan. 1994.
14. J. J. Jacquier, ETSI-European Telecommunications Standards Institute, *Technische Mitteilungen PTT, Switzerland*, July 1990.
15. G. Robin, The European Perspective for Telecommunication Standards, *IEEE Commun. Mag.*, **32**, No. 4, Jan. 1994.

3

SUBSCRIBER SIGNALING

The vast majority of the customers of telecommunication networks are subscribers who are attached to their local exchanges by analog subscriber lines. The signaling between subscriber and local exchange is known as *subscriber signaling* [1–4]. The original, and still predominant, application of subscriber signaling is *plain old telephone service* (POTS) calling. However, subscriber signaling today also supports supplementary services such as *call waiting, call forwarding*, and *caller identification*.

3.1 BASIC SUBSCRIBER SIGNALING

3.1.1 Signaling for an Intraexchange Call

Figure 3.1-1 shows the signaling for an intraexchange call between subscribers S_1 and S_2. The directory number of called subscriber S_2 is 347-9654.

Calling subscriber S_1 starts by going *off-hook* (lifting the handset of the telephone from its cradle). The off-hook is interpreted by the exchange as a *request-for-service* (a call origination, or the activation/deactivation of a subscriber service). In response, the exchange returns *dial-tone*, indicating that it is ready to receive digits. Subscriber S_1 then sends the digits of the called number, using the dial or the keypad of the telephone. After receipt of 3-4-7, the exchange recognizes one of its exchange codes and thus knows that it is the destination exchange for the call.

The exchange can identify the called subscriber S_2 after receipt of the complete called number and checks whether S_2 is free. Assuming that this is the case, it sends a *ringing signal* to alert S_2 and informs S_1 about the call progress with a *ringing-tone*.

When S_2 goes off-hook, an *answer* signal is generated. The exchange then *cuts through* (sets up a path in its switchblock between the subscriber lines). The conversation starts, and the exchange begins to charge S_1 for the call. At the end of the call,

Signaling in Telecommunication Networks, Second Edition, by John G. van Bosse and Fabrizio U. Devetak

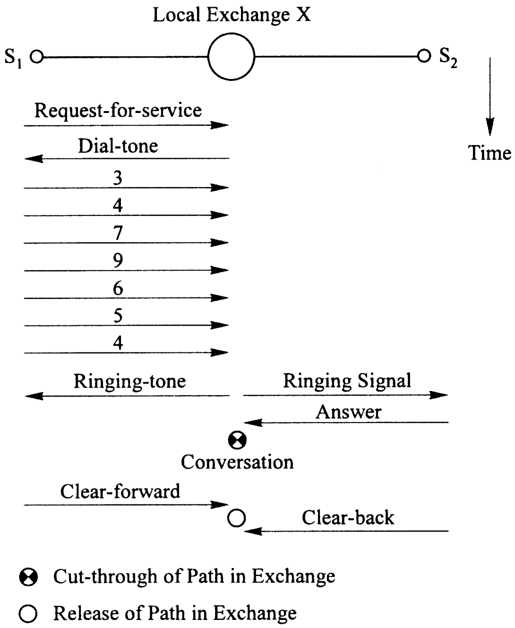


Figure 3.1-1. Signaling for an intraexchange call.

the subscribers put the handsets back in the cradles of the telephones. The signals generated by these actions from the calling and called subscribers are known as, respectively, the *clear-forward* and the *clear-back* signal.

Calling Party Control. The release of connections is usually under control of the calling party. In Fig. 3.1-1, the calling party clears first, and the exchange immediately releases the connection.

If the called party clears first, the exchange starts a timer of, say, 30–60 seconds and releases the call when it receives a clear-forward or on expiration of the timer, whatever occurs first. This is for the convenience of the called party, who may have picked up a phone in one room but wants to have the conversation in another room. The called party can then hang up the first phone, move to the other room and pick up the other phone, while the connection stays up.

Forward and Backward Signals. Call-control signals are categorized as of *forward* and *backward* signals. Forward signals are sent in the direction in which the call is set up (from S_1 to S_2), and backward signals are sent in the opposite direction. The request-for-service signal and the digits of the called number are examples of forward signals. Dial-tone, ringing-tone, and answer are backward signals.

3.1.2 Groups of Subscriber Signals

The signals in the example can be divided into the following four groups.

Supervision Signals. These signals are also called *line signals*. They are sent by subscribers to local exchanges. The forward supervision signals (request-for-service, disconnect by calling party) request the start or end of a connection. The backward supervision signals (answer, disconnect by called party) change the state of a call.

Address Signals. These signals are also known as *digits* or *selection signals*. They are forward signals that are sent by the calling subscriber when dialing the called party number.

Ringing. This is a forward signal sent by the exchange to the called subscriber, to indicate the arrival of a call.

Tones and Announcements. These are audible backward signals (dial-tone, ringing-tone, busy-tone, etc.) sent by an exchange to the calling subscriber, indicating the progress of a call.

3.2 SIGNALING COMPONENTS IN TELEPHONES

This section presents an overview of the circuitry in a telephone, focusing on the components for subscriber signaling.

3.2.1 Telephone

The major components in a telephone are shown in Fig. 3.2-1. The telephone is connected to a line circuit LC in the local exchange by a subscriber line that transfers the subscriber's speech and the subscriber signaling.

Transmitter (microphone) (TR) and receiver (RCV) convert acoustic speech signals to electrical analog signals, and vice versa. Transformer (T) and resistor (R) are part of the speech circuit.

The signaling functions of a telephone are: the generation (controlled by the subscriber) of supervision signals and digits; the conversion of received electrical tone and announcement signals into acoustic signals, and the conversion of the electrical ringing signal into a high-level acoustic signal that can be heard at some distance from the telephone.

3.2.2 Supervision Signals

A telephone can be in two supervision states. When the telephone is not in use, its handset rests in a cradle and depresses the cradle switch (CS)—see Fig. 3.2-1.

In this state, the telephone is *on-hook* (this term has been carried over from the time when the receivers of telephones were resting on a hook). When the telephone is on-hook, switch (CS) connects ringer (RR), in series with capacitor (C), to the subscriber line.

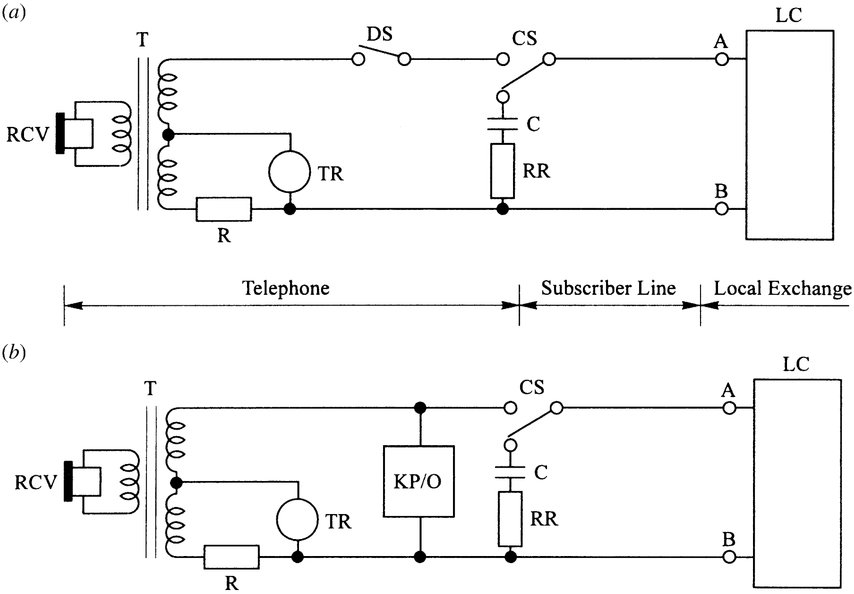


Figure 3.2-1. Components in a telephone, shown in the on-hook state: (*a*) dial telephone and (*b*) keypad telephone. C, Capacitor; KP/O, keypad and oscillator; CS, cradle switch; DS, dial switch; LC, line circuit; RCV, receiver; TR, transmitter; R, resistor; T, transformer; RR, ringer.

When a subscriber starts to use the telephone, he lifts the handset out of its cradle. In this state, which is known as *off-hook*, switch (CS) connects transmitter (TR) and receiver (RCV) to the subscriber line.

When the telephone is off-hook, direct current can flow in the subscriber line. When the telephone is on-hook, capacitor (C) blocks direct current. At the exchange, line circuit (LC) determines the supervision state of the telephone from the presence or absence of direct current in the line.

3.2.3 Address Signals

Address signaling takes place while the telephone is off-hook (CS closed). There are two types of address signals (digits). Today's local exchanges handle both signal types.

Dial-Pulse (DP) Address Signals. In early telephones, the address signals were generated by dials [1,2]. The dial switch (DS) in Fig. 3.2-1(*a*) is linked mechanically to the dial. When the dial is at rest, DS is closed, and the telephone presents a path for direct current (dc) between points A and B. When the dial—after having been rotated by the subscriber—spins back to its rest position, DS opens and closes a number of times, producing a string of *breaks* in the dc path. The number of breaks in a string represents the value of the digit: one break for value 1, two breaks for value 2, . . . , ten

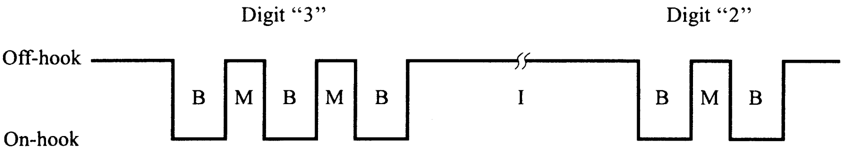


Figure 3.2-2. Dial-pulse address signals. B, Break (60 ms); M, make (40 ms); I, interdigital interval (>300 ms).

breaks for value 0. The nominal length of a break is 60 ms (Fig. 3.2-2). The breaks in a string are separated by *make intervals* of nominally 40 ms. Consecutive digits are separated by an *interdigital interval* of at least some 300 ms.

Dual-Tone Multifrequency (DTMF) Address Signals. Around 1960, it became practical to place transistor oscillators in telephone sets, and this led to the development of DTMF address signaling [3–5]. Figure 3.2-1(*b*) shows a DTMF telephone, which includes a keypad (KP) that controls a dual-tone oscillator (O).

When a subscriber depresses one of the keys on the keypad (KP), oscillator (O) produces two simultaneous tones. A digit is represented by a particular combination of two frequencies: one selected from a *low* group (697, 770, 852, 941 Hz) and the other selected from a *high* group (1209, 1336, 1477, 1633 Hz). This allows 16 digit values, but only 12 of these are implemented on the keypads: digit values 1, 2, . . . , 0, and the special values $*$ and #.

The DTMF frequency combinations have been standardized by ITU-T [6]:

Digit Value	Frequencies (Hz)
1	697 and 1209
2	697 and 1336
3	697 and 1477
4	770 and 1209
5	770 and 1336
6	770 and 1447
7	852 and 1209
8	852 and 1336
9	852 and 1477
*	941 and 1209
0	941 and 1336
#	941 and 1477

3.2.4 Ringing Signal

When the telephone is on-hook (Fig. 3.2-1) and the exchange sends an electrical ringing signal (an alternating current), ringer (RR) produces an audible signal that can be heard in the vicinity of the telephone.

In early telephones, the ringers were electromechanical devices. Modern telephones have electronic ringers.

3.2.5 Tones and Recorded Announcements

These signals have the same electrical characteristics as the speech received during a call. Like speech, they are converted into acoustic signals by receiver (RCV).

3.3 SIGNALING EQUIPMENT AT THE LOCAL EXCHANGE

This section gives an example of the equipment for subscriber signaling at local exchanges. We consider a local SPC (*stored-program controlled*) exchange with a digital switchblock (see Figs. 1.7-4 and 1.7-5).

3.3.1 Overview

Figure 3.3-1 shows the local exchange and a number of subscriber lines. The lines are two-wire bidirectional analog circuits (c). They pass through their line circuits (LC) and are converted into four-wire analog circuits (d) by hybrids (H). First-order PCM multiplexes convert m of these circuits ($m = 24$ or 30) into a four-wire digital multiplex circuit (b) that carries PCM-coded speech (Section 1.5.1). These circuits are attached to digital multiplexer ports (DMP) of the switchblock. The bit format on (b) is as shown in Fig. 1.5-4. Also attached to DMP ports are a number of multiplexed digital *service circuits*.

The switchblock provides temporary 64-kb/s digital paths between a service circuit and a subscriber line—more exactly, a PCM multiplex channel (b) associated

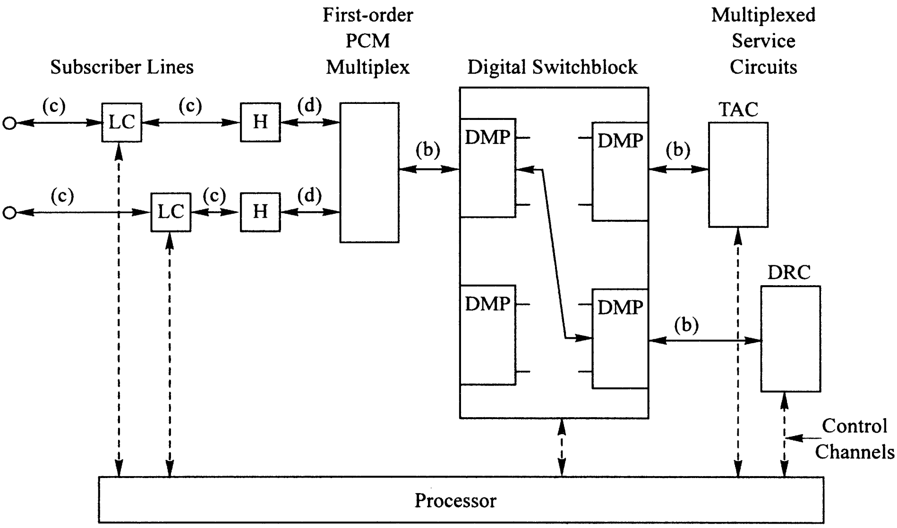


Figure 3.3-1. Equipment for subscriber signaling at a local exchange with a digital switchblock. LC, Line circuits; H, hybrid circuits; DMP, digital multiplexer port; TAC, tone and announcement circuits; DRC, DTMF-digit receiver circuits.

with the subscriber line. These paths transfer PCM-coded tones, DTMF frequencies, and announcements.

The switchblock and the line and service circuits have control channels (CCs) to the exchange processor. This enables the processor to send commands to, and receive information from, these entities.

The implementation of subscriber-signaling functions is manufacturer-specific. In this example, we assume that the line circuits receive the supervision signals and dial-pulse digits from their lines and send ringing signals to the lines.

We also assume that there are two types of service circuits. *Tone and announcement circuits* (TACs) have memories that store PCM sequences for all tones and announcements that can be sent to a subscriber. When, for example, a busy-tone has to be sent to a subscriber line, a switchblock path is set up between an available TAC and the PCM channel associated with the line. The processor then orders the circuit to send a busy-tone. The second type of service circuits are DTMF-digit receiver circuits (DRC). These circuits can provide dial-tone and receive DTMF digits.

3.3.2 Reception of Supervision Signals

Figure 3.3-2 shows a line circuit (LC) in some detail. The circuit can be in two states, which are changed on command from the processor. In the figure, LC is in its "normal" state. Switch (S) connects transformer (T) to the subscriber line. When the telephone is involved on a call, the transformer transfers the (analog) speech between the subscribers. The LC is set to the ringing state when the telephone has to receive a ringing signal.

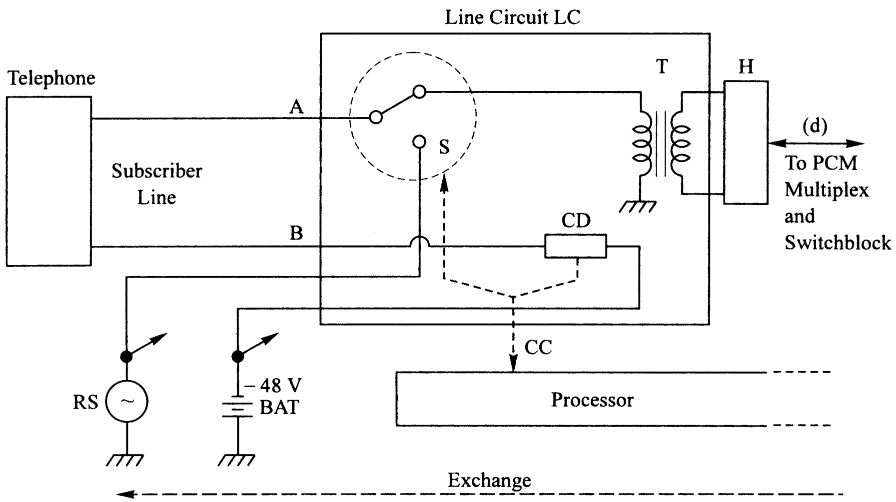


Figure 3.3-2. Components in line circuit, shown in the normal state. T, Transformer; H, hybrid circuit; CD, current detector; CC, control channel; RS, ringing source; BAT, exchange battery; S, switch.

Hook Status. In both LC states, the "hook" state of the telephone is monitored by current detector (CD) and reported to the exchange processor. When the telephone is on-hook (idle), there is no path for direct current between points A and B, and no current flows through CD, which then indicates to the processor that the telephone is on-hook. When the telephone is off-hook, there is a path for direct current between points A and B, and a current flows from ground—through transformer (T), the external path (A–B), and current detector (CD)—to the common battery (BAT). The CD then indicates that the telephone is off-hook.

Call States. The exchange processor keeps track of the "call" state of the telephones and stores these states in its temporary memory. We distinguish the following major call states:

Idle: Not involved in a call

Dialling: Before and during the sending of sending address signals

Calling: Involved in a call as calling party, after having sent the address signals

Ringing: Receiving the ringing signal from the exchange (which indicates an incoming call)

Called: Involved in a call as a called party, and having answered

Determination of Signal Type. A change in the hook status of a telephone is a supervision signal. The processor determines the signal type, based on the present call state, and the type of hook-status change:

Present Call State	Change in Hook Status	Supervision Signal
Idle	To off-hook	Request for service
Dialing		(see Section 3.3.4)
Calling	To on-hook	Clear-forward
Ringing	To off-hook	Answer
Called	To on-hook	Clear-back

Recognition Time. Electrical disturbances on a subscriber line can result in brief off-hook (on-hook) pulses on lines that are on-hook (off-hook). The processor therefore takes no action until the new hook state has persisted for a predetermined recognition time (on the order of 20–40 ms).

Hookswitch Flash (or Flash). This is a fifth supervision signal, sent by a subscriber who is in the calling or called state, to request an action from the local exchange.

The subscriber generates the flash by momentarily depressing the button of cradle switch (CS) (Fig. 3.2-1). This results in a temporary on-hook condition of the telephone. The length of a flash varies widely. Exchanges usually interpret on-hooks of 0.1–1.0 s as flashes and consider longer on-hooks as clear-forward or clear-back signals. The uses of flash signals are discussed later in this section.

3.3.3 Reception of Address Signals

On receipt of a request-for-service from a subscriber line, the processor marks the line as "dialing," selects an idle digit receiver (DRC), orders the switchblock to set up a path between DRC and the line, and commands the DRC to send dial-tone.

If the calling subscriber is using a telephone with dial-pulse address signaling, she rotates the dial, and this generates the digits as strings of "break" and "make" pulses that are detected by current detector (CD) in the line circuit (Fig. 3.3-2) and reported to the processor. On receipt of the first break, the path between the line and DRC is released.

If the calling subscriber is using a telephone with DTMF signaling, she depresses the keys on the keypad. This generates DTMF digits that are received by DRC and reported to the processor. On receipt of the first digit, the dial-tone is turned off. The path between the subscriber line and the DRC is released when the complete called number has been received.

Digit receivers have frequency-selective circuits that are tuned to the individual DTMF frequencies and detect the presence of these frequencies on the subscriber line. The receivers accept a digit only if one frequency of the low group and one frequency of the high group are present simultaneously for at least 70 ms.

Digit Imitation. When a key on a keypad telephone is depressed, the transmitter is disabled. However, there are intervals (between digits) during DTMF signaling when no key is depressed. During these intervals, the transmitter is enabled and may pick up speech, music, or noise in the vicinity of the calling subscriber. These sounds should not imitate DTMF digits. The DTMF frequencies have been chosen to minimize digit imitation, by making DTMF tone pairs distinguishable from naturally occurring sounds [5].

Naturally occurring sounds contain tone pairs whose frequency ratios are "simple" fractions (such as 1:2, 3:5, 2:3, 3:4, 4:5).

Suppose now that a sound has a 1336-Hz component. This is detected in the digit receiver by the frequency-selective circuit tuned to this frequency. The most likely companion frequencies in this sound are listed below:

Frequency (Hz)	Adjacent DTMF Low-Group Frequencies (Hz)
(1:2) × 1336 = 668	—, 697
(3:5) × 1336 = 801	770, 852
(2:3) × 1336 = 891	852, 941
(3:4) × 1336 = 1002	941, —
(4:5) × 1336 = 1069	941, —

Note that these companion frequencies are about midway between two adjacent low-group DTMF frequencies. Therefore, none of the selective circuits tuned to the low-group DTMF frequencies detects a signal, and the natural sound is not accepted as a digit. The same happens with natural sounds that contain one of the other DTMF frequencies.

3.3.4 Ringing Signals

The ringing signal is a high-level 20–25 Hz signal analog, of typically 100 V rms, and designed to drive the electromechanical ringers in the early telephones. This ringing signal cannot be provided by a service circuit, because there are no PCM codes to represent these high-level voltages. The ringing signal is therefore injected into the line circuit (Fig. 3.3-2).

When a subscriber line has to be rung, the processor sets the line circuit to the ringing state. In this state, switch (S) connects a common ringing-voltage source (RS) to the A-wire of the subscriber line. The alternating ringing current passes through the external circuit (subscriber line, capacitor (C), and ringer (RR)—see Fig. 3.2-1) and detector (CD). However, this current is not large enough to cause CD to indicate "off-hook."

When the called telephone goes off-hook, the ringer current and the direct current from battery (BAT) flow through CD. An off-hook is reported to the processor, which then changes the LC state to normal.

3.4 TONES, ANNOUNCEMENTS, AND RINGING

A calling subscriber receives tones or announcements that inform him about the progress of the setup of his call. These tones originate at the local exchange of the caller, at intermediate exchanges along the connection, or at the local exchange of the called party (Fig. 3.4-1).

3.4.1 Tone Formats

In the early years of telecommunications, all call-progress signals were tones, at frequencies of 400–600 Hz. These tones, some of which are still in use, are either continuous or repeating "on–off" cycles with a certain cadence. The frequencies and cadences were established by the telecoms of individual countries, well before the beginnings of international standards. The meanings of tones, and their frequencies and cadences, are therefore somewhat country-specific.

3.4.2 Basic Tones

The following functional tone signals are used in all countries.

Dial-Tone. This is sent by the local exchange serving the calling subscriber, to indicate that the exchange is ready to receive the called number. In most countries, dial-tone is continuous.

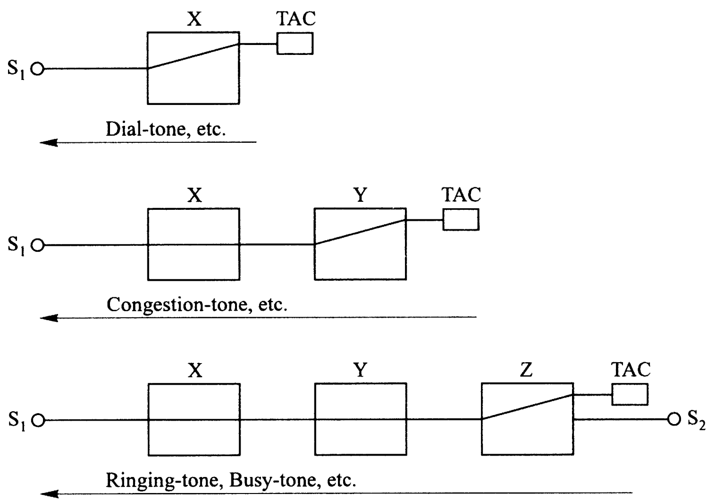


Figure 3.4-1. Call-progress tones and announcements. X, Local exchange of calling party; Y, intermediate exchange; Z, local exchange of called party; TAC, tone and announcement circuit.

Ringing-Tone (Audible Ring Tone). This is sent by the local exchange of the called subscriber to indicate that the subscriber is being alerted. In most countries, ringing-tone has a cadence, for example:

On (seconds)	Off (seconds)	Country
2	4	United States, Canada, . . .
1	4	Brazil, Mexico, . . .

The United Kingdom, Australia, and other countries that follow British telecommunication practices have the following ringing-tone cadence:

On (seconds)	Off (seconds)	On (seconds)	Off (seconds)
0.4	0.2	0.4	2

The ringing signals have the same cadences as the ringing-tones.

Busy-Tone. This is sent by the local exchange that serves the called subscriber when that subscriber is busy. Examples are as follows:

On (seconds)	Off (seconds)	Country
0.5	0.5	United States, Canada, . . .
0.25	0.25	Brazil, Mexico, . . .

3.4.3 Tones and Announcements in Other Failed Setups

Tones and announcements to indicate that a setup has failed for a reason other than "called party busy" fall into the following broad categories.

Congestion (All Trunks Busy, ATB). The call setup fails because it has reached an exchange at the time that the exchange has no available trunks in its route set to the called exchange.

Invalid or Nonworking Called Number. An exchange along the route determines that the received called number contains invalid information (say, an invalid area or exchange code), or a subscriber number that is presently not allocated.

In the early days of telephony, busy-tone was the universal signal for all setup failures. However, it is important to give the calling party more detailed information, so that she can decide whether and when to repeat the call. For example, if the called subscriber is busy, it makes sense to repeat the call attempt after a few minutes. However, if the called number is invalid, all succeeding attempts are doomed to failure. Again, the tones and announcements for these failure cases are country-dependent. A few examples are outlined below.

Number Unobtainable Tone. This is used primarily in the United Kingdom and in countries that follow British telecommunication practices. It indicates that the called number is not a working number. In the United Kingdom, the tone is continuous. Other countries use cadences, for example:

On (seconds)	Off (seconds)	Country
2.5	0.5	Australia, South Africa, . . .

Congestion-Tone. This is sent by exchanges when the call cannot be set up because of congestion in the network (no outgoing trunk available). In the United States, this tone is known as "reorder" tone. Some typical cadences are:

On (seconds)	Off (seconds)	Country
0.25	0.25	United States, Canada, . . .
0.5	0.5	Belgium, The Netherlands, . . .

In a number of countries (Austria, Brazil, Italy, Russia, etc.), congestion is still indicated by busy-tone.

Special Information Tone (SIT). This is a cadenced sequence of three frequencies, as shown in Fig. 3.4-2. It is used in many European countries to indicate that the received called number is not a working number.

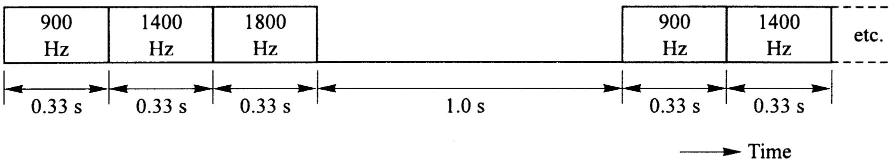


Figure 3.4-2. Special information tone.

SIT and Recorded Announcement. In the United States, one cycle of SIT is often sent before an announcement that gives more details about the problem, for example: "Your call cannot be completed as dialed. . . ." In case the called number of a subscriber has been changed, the new number is usually included in the announcement.

An extensive survey of tones and recorded announcements used in various countries can be found in [7].

3.5 SUBSCRIBER SIGNALING FOR SUPPLEMENTARY SERVICES

Up to this point, we have discussed subscriber signaling for POTS (plain old telephone service) call control. Today, customers in many countries can subscribe to *supplementary services*, for which they are charged a monthly fee. Every local exchange has a database with entries for each subscriber, listing the supplementary services available to that subscriber.

The supplementary services offered in the United States can be divided into two groups. *Custom calling* services became available in the late 1960s, with the introduction of *stored-program controlled* local exchanges in the network. *Custom local area signaling services* (CLASS) have since been introduced in the United States.

This section examines some of these services and outlines the additions to subscriber signaling to support them.

3.5.1 Custom Calling Services

Custom calling services do not require special signaling hardware in telephone sets. Subscribers have to invoke some of these services by dialing special digit sequences called *feature access codes*. These codes are distinguishable from called numbers because their initial digits include a * (asterisk) or a # (pound sign) when dialed from DTMF telephones, or special digit sequences (e.g., 11) when dialed from dial-pulse telephones. Some services require the subscriber to send hookswitch flashes during the call.

The local exchanges on which these services are offered have to be equipped with tone and announcement circuits (TACs) that can provide a number of additional audible signals and messages to the subscriber.

Some widely used calling services are outlined below [3].

Call Waiting Service. Suppose that subscriber S_1 is marked at his local exchange for *call waiting service*. When S_1 is engaged in a call with subscriber S_2 and is called by another subscriber S_3, he is alerted by his local exchange with a *call waiting* tone. By sending a hookswitch flash, S_1 can put subscriber S_2 on hold and be connected to S_3. From this point on, S_1 can alternate between S_2 and S_3 with subsequent flashes. Call waiting tone in the United States is a 440-Hz tone with a cadence of 0.3 s on–10 s off.

Call Forwarding Service. When S_1 is marked at the local exchange for *call forwarding service*, he can activate this service by dialing a *feature access* code. The exchange responds with dial-tone, after which S_1 enters the number of subscriber S_2 to whom his incoming calls are to be forwarded. The exchange acknowledges receipt of the number with two short "beeps" and then sets up a call from S_1 to S_2. Subscriber S_1 then hears ringing-tone. When S_2 answers, the call forwarding for S_1 is activated. If S_2 is busy or does not answer, S_1 disconnects and repeats the activation request, which is then accepted by the exchange without setting up a call to S_2. When call forwarding has been activated, S_1 can still make outgoing calls, and his telephone still rings for incoming calls. However, these calls are forwarded until he deactivates the service, by dialing.

Three-Way Calling. When S_1 is registered for *three-way calling service*, he can call S_2 and then add a call to S_3. Having established the call to S_2, subscriber S_1 flashes. The local exchange returns dial tone, and S_1 enters S_3's number. The exchange then sets up the call to S_3 while maintaining the call to S_2. If S_3 is busy or does not answer, S_1 can end the new call (but remain in conversation with S_2), by flashing twice.

3.5.2 Custom Local Area Subscriber Services

CLASS services are new supplementary services that have been introduced in the United States [8]. They depend on the availability of common-channel signaling system No. 7 (see Chapter 11), which can transfer the calling number from the calling to the called exchange. Some CLASS also require additions to subscriber signaling at the local exchanges, such as recognition of *feature access codes*, TACs that include the required tones and announcements, and distinctive ringing. A few examples are outlined below.

Distinctive Ringing. A subscriber registered for this service is alerted by a ringing signal with a distinctive cadence when called by certain specified lines. The subscriber can list up to ten such calling numbers. To make a change in the list, the subscriber dials a feature access code. This initiates a dialogue with the exchange during which the subscriber is prompted by recorded messages and enters the changes by sending digit strings that represent the special calling numbers. For this service, the local exchanges have to be able to generate the distinctive ringing signal and to send the necessary recorded messages.

Selective Call Rejection. A subscriber registered for this service can enter a list of up to ten numbers from which she does not want to receive calls. Again, the subscriber can change this list by dialing a feature access code and sending the digits of the calling numbers when prompted by messages from the exchange.

Caller ID (Calling Number Delivery). On calls to subscribers registered for this service, the local exchange signals the calling number, and the name of the calling subscriber, to a display device attached to the subscriber's telephone. The number is sent by a specialized type of service circuit in the local exchange.

The digits and characters are sent in a digital format in which 0's and 1's are sent as two voiceband frequencies (frequency-shift keying). The signaling takes place during the silent interval after the first burst of ringing.

A calling subscriber can prevent the presentation of the calling number by prefixing the called number with a feature access code.

3.5.3 Supplementary Services in Other Countries

Supplementary services are country-specific. Some of the services described above are also offered outside the United States, others are not. There are also services that are available in other countries but not in the United States. For example, *wake-up* service allows a subscriber to enter a feature access code and then dial a number of digits that specify the day, and time of day, when he wants to receive a wake-up call. Also, a subscriber can dial feature access codes to activate or deactivate the *do not disturb* service. When the service is activated, all incoming calls to the subscriber are diverted by her local exchange to a recorded announcement that indicates that the subscriber does not want to receive calls.

The feature access codes and signaling procedures to activate/deactivate or modify supplementary services vary from country to country. Many countries offer supplementary services to subscribers with DTMF telephones only and use feature access codes that always include a # and/or * [9].

3.6 OTHER APPLICATIONS OF DTMF SIGNALING

DTMF address signaling was designed as a convenience feature for subscribers, providing a faster and more convenient way to send called numbers to their local exchanges.

Another aspect of DTMF signaling, which was not considered at the time, is that the frequency pairs of DTMF digits are in the voiceband range (300–3400 Hz) and enable a calling subscriber to send DTMF signal to other exchanges along the connection. This is not possible with dial-pulse signaling because most of the energy of dial-pulse signals is concentrated at frequencies below 300 Hz, and these signals cannot be transferred reliably across the network.

Some applications of the transfer of DTMF digits between the parties in a connection are outlined below.

3.6.1 Caller Interaction

Private branch exchanges (PBXs) can be equipped with *caller-interaction service circuits* (CISCs), which automate the handling of incoming calls. A CISC stores a number of spoken messages that are played on command of the PBX processor and has a DTMF digit receiver. This enables the PBX to "interact" with the caller. When a call arrives at the PBX, the incoming trunk is first connected to a CISC and plays an announcement that prompts the caller to send one or more digits, for example:

> Thank you for calling XYZ airlines. If you are calling from a pushbutton telephone, please press 2 to purchase tickets; press 3 to change your reservation; press 4 for flight arrival and departure information, ... If you are calling from a dial telephone, stay on line, and you will be connected when an agent becomes available.

If the caller is using a pushbutton phone, he then sends the appropriate DTMF digit, and the PBX processor sets up a connection to an agent who is qualified to handle the requested service. This arrangement has become quite popular with business organizations, because it reduces the need for PBX attendants. Calling subscribers are less enthusiastic, especially when having to respond to a series of these prompts.

Telecoms are also equipping their local exchanges for caller interactions. This opens the door to *intelligent network* services, which are described in Chapter 17.

3.6.2 Signaling to Competing Long-Distance Carriers

A DTMF application that was not foreseen by the Bell System (the main developer and promoter of DTMF signaling) opened the door to competition in long-distance telecommunications in the United States. An early form of this competitive service is shown in Fig. 3.6-1. A *competing carrier* (CC) maintained PBXs in several cities, which were connected by subscriber lines to the respective local exchanges and were interconnected by intercity trunk groups TG_{CC}, owned by CC.

A subscriber (S_1) with a DTMF telephone, and located in city A, could call subscriber S_2 in city B by making a local call to CC's PBX_A in city A. When the connection was established, PBX_A would prompt S_1 to send—with DTMF digits—the called and calling numbers (the calling number was needed for charging). After this, PBX_A would extend the call to PBX_B in city B, on a trunk of TG_{CC}. The call would then be completed by PBX_B, as a local call in city B.

Dialing was cumbersome, and speech transmission was often substandard, because the connection involved four subscriber lines (SLs) instead of two. However, the cost savings attracted a number of customers, and this allowed CC companies to become established as "alternative" long-distance carriers.

The operators of the local telecommunication networks in the United States are now required to provide *equal access* to all interexchange carriers, and the connections shown in Fig. 3.6-1 are no longer needed.

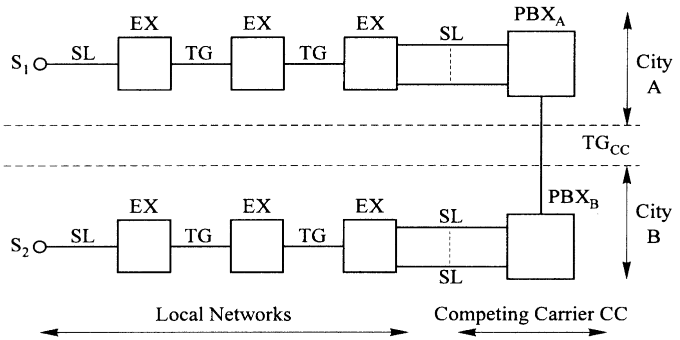


Figure 3.6-1. Original configuration for long-distance calls via competing carrier (CC). SL, Subscriber lines; TG, trunk groups in local networks of cities A and B; TG_{CC}, group of tie trunks owned by CC; $PBX_{A,B}$, private branch exchanges owned by CC; EX, exchanges in the local networks of cities A and B.

3.7 DIALING PLANS

A dialing plan at a local exchange is a list of digit strings that can be dialed by its subscribers. A U.S. dialing plan consists of a *public office dialing plan* (PODP) and possibly a number of *customized dialing plans* (CDPs). Dialing plans differ from numbering plans in the sense that they include all the prefixes that must be dialed before the digits of the numbering plan, as well as special codes that are controlled by each individual telecom to provide supplementary services. International numbering plans and the North American numbering plan (NANP) are described in Section 1.2.

3.7.1 Public Office Dialing Plan

PODP lists the digit strings that can be dialed by the general public. An example of a PODP is shown in Table 3.7-1 (see also Section 1.2.1). It covers called party numbers and service access codes.

TABLE 3.7-1 Public Office Dialing Plan (U.S.)

Type of Number	Dialing Pattern
(a) Intra-NPA call	EC(3) + LN(4)
(b) Inter-NPA call	1 + AC(3) + EC(3) + LN(4)
(c) International call	011 + (7 to 15 digits)
(d) Operator-assisted national call	0 + AC(3) + EC(3) + LN(4)
(e) Operator-assisted international call	01 + (7 to 15 digits)
(f) LEC operator	0
(g) IXC operator	00
(h) Special numbers	*N*11
(i) Inter-LATA call with specified carrier	101 + CIC + 1 + AC(3) + SN(7)
(j) Service access codes (DTMF phones)	**XX* or 7*X*#
(k) Service access codes (dial-pulse phones)	11*XX* or 7*X*

Called Party Numbers. A calling subscriber dials a subscriber number (a), a national number (b, d), or an international number (c, e), depending on whether the called party is in the same numbering plan area (NPA) as the calling party, or in a different NPA, or in a different country altogether. On national and international calls, the subscriber can request operator assistance (d, e). A subscriber can also request operator assistance without dialing a called number (f, g).

If the called party is located outside the LATA of the caller, the connection involves an *interexchange carrier* (IC or IXC). Each subscriber has a preassigned IXC of his choice, known by the local exchange. On inter-LATA calls, the local exchange routes the call to an exchange of the preassigned IXC, unless the caller has dialed a carrier access code (CAC) as in (i). After the CAC the subscriber dials as in (b). The CAC has the following format:

$$\text{CAC}(7) = 101 + \text{CIC}(4)$$

where CIC(4) $= XXXX$ (Section 1.2).

Dialing the CAC (i) bypasses the preassigned IXC and routes the call through the IXC identified by the carrier identification code (CIC).

Prefixes. To expedite call processing, an exchange must recognize the type of dialed number as soon as possible and that is achieved with prefixes. In the United States, all prefixes start with a 0 or a 1. Since called numbers always start with $N > 1$, the exchange recognizes a prefix on receipt of the first digit.

Subscriber numbers (a) and $N11$ codes (h) are dialed without prefix. They become distinguishable on receipt of the third digit.

National numbers (b, d) are prefixed by 1 (no operator assistance) or 0 (operator assistance). International numbers (c, e) have a 011 or 01 prefix. The exchange thus knows the called number type, and the need for operator assistance, after at most three digits.

The CAC dialing pattern is recognized on the second digit of the prefix; the 1 in the third position is needed to grandfather older three-digit CICs.

Timeouts. Some dialed patterns, for example, 0 (call to a local carrier's operator), require a timeout for recognition. If a digit (other than another 0) is received before the expiration of the timer, the exchange concludes that the caller is dialing an operator-assisted national number. If the timer expires and no additional digits have been received, the exchange concludes the caller is requesting a LEC operator.

Feature Access Codes. These codes are dialed to activate or deactivate a supplementary service feature. Their formats are $11XX$ or $7X$ (from dial-pulse telephones) and $*XX$ or $7X\#$ (from DTMF phones). For example:

72 or 72#	Activate call forwarding
73 or 73#	Cancel call forwarding

1170 or *70	Deactivate call waiting for one call
11676 or *67	Deactivate caller ID for one call

3.7.2 Customized Dialing Plans

At the request of a multilocation business, a telecom can provide a CDP for calls between employees of the organization. A typical customized plan allows the caller to dial an access code followed by a short number. The access code indicates whether the subsequent digits belong to the CDP or to the PODP. Usually access code 8 indicates a CDP, and access code 9 indicates the PODP.

3.8 ACRONYMS

ATB	All trunks busy
BAT	Exchange battery (48 V)
C	Capacitor
CAC	Carrier access code
CC	Control channel, competing carrier
CD	Current detector
CDP	Customized dialing plan
CIC	Carrier identification code
CISC	Caller-interaction service circuit
CLASS	Custom local area signaling services
CS	Cradle switch
DP	Dial pulse
DMP	Digital multiplexer port
DRC	DTMF-digit receiver circuit
DS	Dial switch
DTMF	Dual-tone multifrequency
H	Hybrid circuit
IC	Interexchange carrier
IXC	Interexchange carrier
KP/O	Keypad and DTMF oscillator
LATA	Local area transport and access
LC	Line circuit
LEC	Local exchange carrier
NANP	North American numbering plan
NPA	Numbering plan area
PBX	Private branch exchange
PODP	Public office dialing plan
POTS	Plain old telephone service
R	Resistor
RAC	Recorded announcement circuit
RCV	Receiver (earpiece of telephone)

RR	Ringer
RS	Ringing source
S	Subscriber, switch
SC	Service circuit
SIT	Special information tone
SL	Subscriber line
SPC	Stoved program controlled
T	Transformer
TAC	Tone and announcement circuit
TG	Trunk group
TR	Transmitter (mouthpiece of the telephone)

3.9 REFERENCES

1. S. Welch, *Signalling in Telecommunications Networks*, Peter Peregrinus Ltd., Stevenage, UK, 1981.
2. R. L. Freeman, *Telecommunication System Engineering*, 2nd ed., Wiley–Interscience, Hoboken, NJ, 1989.
3. *Engineering and Operations in the Bell System*, 2nd ed., AT&T Bell Laboratories, Murray Hill, NJ, 1983.
4. *A History of Engineering and Science in the Bell System*, Bell Telephone Laboratories, 1982.
5. L. Schenker, Pushbutton Calling with a Two-Group Voice-Frequency Code, *Bell Syst. Tech. J.*, **38**, 1960.
6. *General Recommendation on Telephone Switching and Signalling*, Recommendations Q.23–Q.24, ITU-T, Geneva, 1998.
7. *International Telephone Service Operation*, Recommendations E.180–E.183, and Suppl. Nos. 2 and 4, ITU-T, Geneva, 1988–1998.
8. S. M. Boyles, R. L. Corn, and L. R. Moseley, Common Channel Signaling: The Nexus of an Advanced Communications Network, *IEEE Commun. Mag.*, **28**, July 1990.
9. *International Telephone Service Operation*, Rec. E.131–E.132, ITU-T, Geneva, 1988.

4

CHANNEL-ASSOCIATED INTEREXCHANGE SIGNALING

Channel-associated interexchange signaling (CAS)—also known as *per-trunk signaling*—has been in existence from the beginning of automatic telephony and was the only form of interexchange signaling until 1976. It is still used in telecommunication networks but in many countries has largely been replaced by *common-channel signaling.*

Early CAS systems were developed independently by individual equipment manufacturers and exist in many varieties. Later CAS systems, notably those developed after the Second World War, show the increasing influence of national and international standards.

This section describes three important CAS systems and their use on *frequency-division multiplexing* (FDM) analog trunks and *time-division multiplexing* (TDM) digital trunks (Sections 1.4.5 and 1.5.2).

The acronyms in the figures of this chapter are explained in Section 4.5.

4.1 INTRODUCTION

4.1.1 Interexchange Signaling Example

Figure 4.1-1 shows a typical interexchange signaling sequence for a call from subscriber S_1 to subscriber S_2. The subscriber signaling for the call is not shown.

After exchange A has received the called number from S_1, it decides to route the call via intermediate exchange B. Exchange A seizes an available trunk T_1 and sends a *seizure* signal on the trunk. Exchange B responds with a *proceed-to-send* (or *wink*) signal, indicating that it is ready to receive the digits of the called number. Exchange A sends the digits and then *cuts through* (sets up a path in its switchblock between the subscriber line of S_1 and T_1).

Signaling in Telecommunication Networks, Second Edition, by John G. van Bosse and Fabrizio U. Devetak

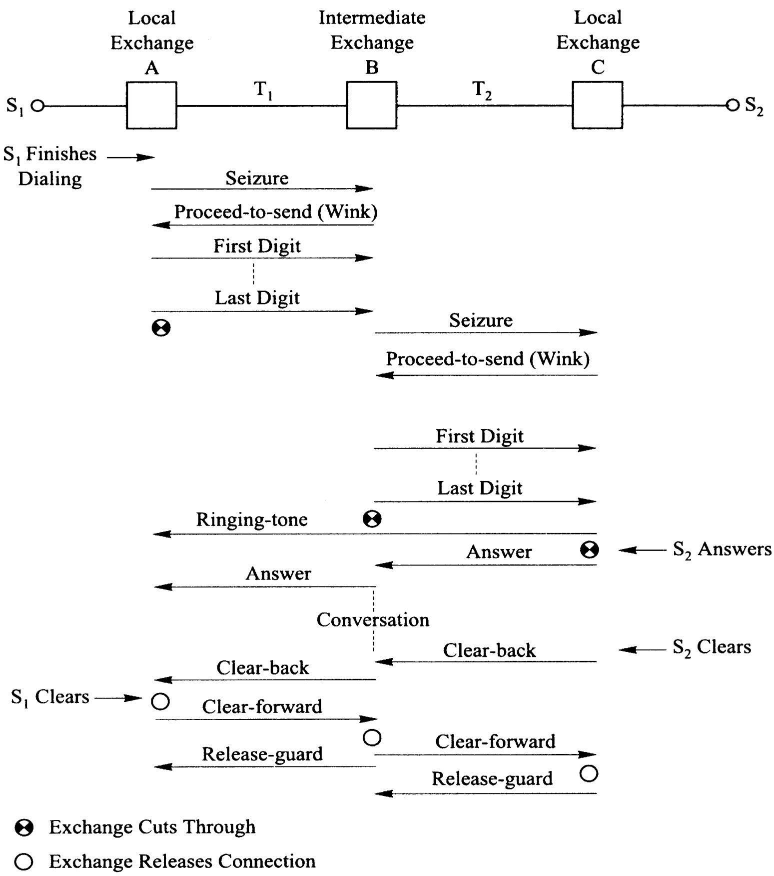


Figure 4.1-1. Interexchange signaling.

When exchange B has received the complete called number, it seizes an available trunk T_2 to destination exchange C and sends a seizure signal on the trunk. Exchange C responds with a wink signal, after which exchange B sends the digits of the called number and cuts through a path between trunks T_1 and T_2.

Exchange C then checks whether called subscriber S_2 is idle. If this is the case, C sends a ringing signal to S_2 and ringing-tone on trunk T_2. Because exchanges A and B have cut through, there is a connection between the calling subscriber S_1 and exchange C, and subscriber S_1 hears a ringing-tone.

When S_2 answers, exchange C cuts through a path between trunk T_2 and subscriber S_2. It also sends an *answer* signal on T_2, and exchange B repeats the signal on trunk T_1. Assuming that originating exchange A is responsible for charging the call, it establishes a billing record that includes the calling and called numbers, the date, and the time of answer.

The conversation now begins. In this example, called party S_2 hangs up first. Exchange C sends a *clear-back* signal to exchange B, which repeats the signal to exchange A. Like intraexchange calls, interexchange calls are usually controlled by the calling party (Section 3.1.1). On receipt of the clear-back, exchange A stops charging and enters the time when it received the clear-back in the billing record of the call. It also starts a 30–60 second timer. It then awaits a clear-forward from calling party S_1, or the expiration of the timer, and initiates the release of the connection when one of these events occurs.

The release takes place in the following way. Exchange A releases its path between S_1 and trunk T_1 and sends a clear-forward signal to exchange B, which releases its path between T_1 and T_2 and repeats the clear-forward to exchange C. This exchange then clears its path between T_2 and called subscriber S_2.

When exchanges B and C have completed the release of, respectively, T_1 and T_2, they send *release-guard* signals (to, respectively, exchanges A and B). When A and B receive the release-guard, they know that they can again seize, respectively, T_1 and T_2 for new calls.

This scenario is typical for CAS in general, but we shall encounter variations that are specific to individual CAS systems.

4.1.2 Groups of Interexchange Signals

Of the four groups of subscriber signals discussed in Section 3.1.2, three also exist in channel-associated interexchange signaling.

Supervision Signals. These signals are also known as *line signals*. They represent events that occur on the trunk, such as seizure, proceed-to-send, answer, or clear-forward. While the majority of supervision signals are used in all CAS systems, there are system-specific differences in the sets of supervision signals.

Address Signals. These signals are also known as *selection signals, digits*, or *register signals*. The digits are used primarily to indicate the called number but can also have other meanings. Like dial telephones, early CAS systems used dial-pulse address signals. The systems described in this chapter have *multifrequency* (MF) address signaling, similar to the DTMF (dual-tone multifrequency) signaling in pushbutton telephones. The MF frequencies are system-specific and different from those used in DTMF.

Tones and Announcements. Examples are ringing-tone and busy-tone. The tones and announcements in interexchange signaling are the same as in subscriber signaling (Section 3.2.5).

4.1.3 CAS Equipment at the Exchanges

We consider a *stored-program controlled* (SPC) intermediate exchange with a digital switchblock (Fig. 4.1-2). The switchblock provides temporary bidirectional 64-kb/s paths. The *digital multiplexed ports* (DMPs) are first-order TDM circuits (b) with frame formats as described in Section 1.5.2.

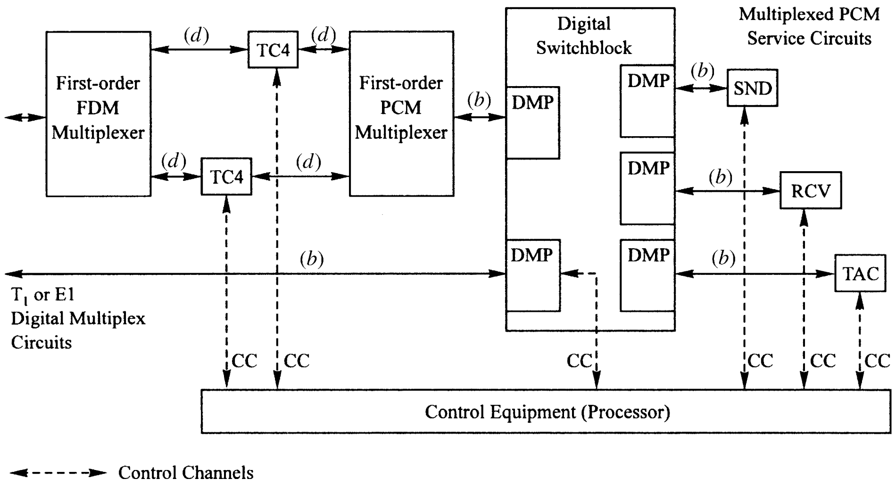


Figure 4.1-2. Exchange equipment for channel-associated interexchange signaling.

All trunks have CAS. The attachment of analog and digital (PCM) trunks to the switchblock is as described in Section 1.7.2. First-order TDM multiplexes carrying PCM trunks (b) are directly connected to DMPs. Analog trunks on first-order FDM transmission systems are first demultiplexed into individual four-wire analog circuits (d), which pass through trunk circuits (TC4) and then enter a first-order PCM multiplexer, where they are converted into TDM digital circuits (b).

Supervision Signaling. The TC4 circuits of analog four-wire trunks inject outgoing supervision signals into the trunk and detect supervision signals received on the trunk. Each TC4 has a control channel (CC). It is used by the processor to order the TC4 to send signals, and by the TC4 to report received signals.

The supervision signaling information for the trunks in first-order digital trunk multiplexes is in certain bits of the bit streams (b). Control channel (CC) to the switchblock has two functions. In the first place, it transfers processor commands to set up and release switchblock paths. In the second place, it is used for communications between the processor and those DMPs that serve groups of multiplexed digital trunks. The processor can order a DMP to send out a supervision signal on a specified trunk in the multiplex, and a DMP reports the supervision signals received from the trunks in its multiplex to the processor.

Service Circuits. Address signaling and the sending of tones/announcements are done with service circuits. The exchange has pools of digital multiplexed service circuits that are attached to DMPs on the switchblock. Tone and announcement circuits (TACs) send tones and announcements. Digit-senders (SND) and digit-receivers (RCV) send and receive MF address signals to and from the trunks. These circuits are signaling system specific. If an exchange has trunks with two CAS signaling systems, it needs separate groups of senders and receivers for each system. The circuits are controlled by the control equipment (processor) of the exchange, via control channels (CCs).

To send address signals or tones/announcements on a trunk, or to receive address signals from a trunk, the processor seizes an available service circuit of the proper type and commands the switchblock to set up a path between the circuit and the trunk. It then orders a SND or TAC circuit to send specific digits or tones, or orders a RCV circuit to report the received digits. When the sending or receiving has been completed, the path is released, and the circuit is returned to its pool.

The numbers of TACs and RCV circuits in an exchange are small compared to the number of CAS trunks on the exchange, on the order of about one circuit of each type for every 10–20 trunks.

4.1.4 Interexchange Call-Control Definitions

It is useful to introduce some terms that are frequently found in descriptions of call-control signaling, with the aid of Fig. 4.1-1.

Outgoing and Incoming. These terms can be applied to trunks and to exchanges. An *outgoing* exchange seizes *outgoing trunks*, sends *forward* signals, and receives *backward* signals, on its outgoing trunks. An *incoming* exchange receives forward signals and sends backward signals, on its *incoming* trunks. When discussing the signaling between exchanges A and B in Fig. 4.1-1, exchange A is the outgoing exchange, and B is the incoming exchange. When describing the signaling between exchanges B and C, exchange B is the outgoing exchange, and C is the incoming exchange.

Trunk T_1 (which is seized by exchange A) is regarded at A as an outgoing trunk and at B as an incoming trunk. In the same way, trunk T_2 is an outgoing trunk at exchange B and an incoming trunk at exchange C. If T_1 is a one-way trunk, it is always an outgoing trunk at A and an incoming trunk at B. However, if T_1 is a bothway trunk, its role at exchanges A and B varies per call, depending on whether it has been seized by A or B. The same holds true for trunk T_2.

Originating and Terminating Exchanges. The *originating* exchange in a call is the local exchange serving the calling subscriber, and the *terminating* (or *destination*) exchange is the local exchange of the called subscriber. In the example of Fig. 4.1-1, exchange A is the originating exchange, and C is the terminating exchange.

Overlap and En-bloc Address Signaling. In the example of Fig. 4.1-1, exchange A seizes trunk T_1 after it has received the complete called number from subscriber S_1. Likewise, exchange B seizes trunk T_2 after receiving the complete number from exchange A. This means that, once the exchanges receive a proceed-to-send, they send out the complete called number in one uninterrupted stream. This mode of address signaling is called *en-bloc* register signaling.

However, exchanges can generally make route decisions after receipt of just the initial part of the called number. For example, if the called number in the example is a subscriber number, consisting of an exchange code followed by a line number (EC-LN), exchange A can seize trunk T_1 after receipt of EC from the calling subscriber and can then send EC to exchange B. This exchange again can seize its outgoing trunk T_2 after receipt of the EC and send the EC to exchange C. Exchanges A and B thus send out the initial digits of the called number while still receiving the later digits of the number, which are sent as soon as they have been received from the calling subscriber, or from the preceding exchange in the connection. This mode of register signaling is called *overlap* register signaling. The decision to use *en-bloc* or overlap address signaling is made by the individual telecoms. Overlap address signaling results in faster call setups.

Link-by-Link and End-to End Signaling. Signaling by two exchanges at the two ends of a trunk is called link-by-link signaling. In Fig. 4.1-1, supervision and address signaling are link-by-link. In general, supervision signaling is always link-by-link. Address signaling is link-by-link in most, but not all, CAS systems.

In end-to-end address signaling, the digit sender in the originating exchange sends address signals successively to digit receivers in the second and later exchanges in the connection.

4.1.5 CAS Signaling Systems

The CAS signaling systems discussed in this chapter are:

Bell System multifrequency (MF) signaling
CCITT No. 5 signaling (only mentioned briefly)
R2 signaling

4.2 BELL SYSTEM MULTIFREQUENCY SIGNALING

This section describes the multifrequency signaling system that was introduced by the Bell System after the Second World War [1–3]. It is still in use today, mostly in local U.S. networks. A nearly identical signaling system, known as the R1 signaling system [4] and defined by ITU-T, is used on international trunk groups in the North American network (e.g., groups between the United States and Canada).

Supervision and address signaling are link-by-link. The system can be used on one-way two-wire trunks, one-way and bothway FDM analog trunks, and one-way and bothway TDM digital trunks (see Sections 1.4.5 and 1.5.2).

The supervision signaling is described in Sections 4.2.1 through 4.2.3. The MF address signaling is discussed in Section 4.2.4.

4.2.1 Supervision Signaling

In Chapter 3 we discussed the on-hook and off-hook states of a telephone. In Bell System MF signaling, we speak of the on-hook (idle) and off-hook (in use) states of a trunk. These states can be different at the exchanges connected by the trunk. Each exchange continuously sends the trunk state at its end to the other exchange. This is known as *continuous two-state* signaling. Changes in trunk state are supervision signals.

Supervision Signals. The repertoire of supervision signals and the corresponding state changes are:

Forward Signals	State Change
Seizure	On-hook to off-hook
Clear-forward	Off-hook to on-hook

Backward Signals	State Change
Answer	On-hook to off-hook
Clear-back	Off-hook to on-hook
Proceed-to-send (wink)	Off-hook pulse (120–290 ms)

Consider a trunk between exchanges A and B. If the trunk is a one-way trunk that can be seized only by exchange A, this exchange is the outgoing exchange for all calls and sends forward signals. Likewise, exchange B is always the incoming exchange and sends backward signals. If the trunk is a bothway trunk, the exchange that seizes the trunk for a call is the outgoing exchange for that call.

Figure 4.2-1 shows the states, and the forward and backward supervision signals, for a typical call. Exchange A is the outgoing exchange.

Initially, the trunk is on-hook at both ends. Exchange A seizes trunk T and sends a forward off-hook (seizure signal). Exchange B then connects a digit receiver to the trunk and sends a (backward) wink signal. After receipt of the wink, exchange A sends the digits of the called number. When the call is answered, exchange B

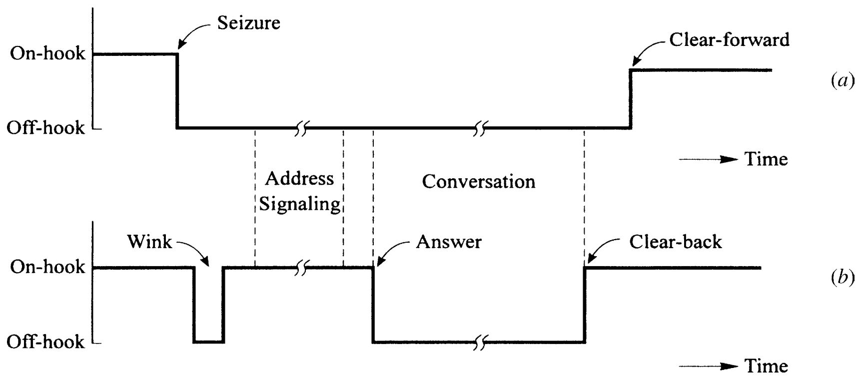


Figure 4.2-1. Supervision signals for a call: (*a*) sent by outgoing exchange A and (*b*) sent by incoming exchange B.

sends an off-hook (answer signal). During the conversation, both exchanges are sending off-hook.

In this example, the called party clears first, and exchange B starts sending on-hook (clear-back signal). When the calling party clears, exchange A releases the trunk and sends on-hook (clear-forward). In response, exchange B clears the trunk at its end.

The signaling system does not include a release-guard signal. Therefore, when outgoing exchange A releases the trunk, it starts a timer that expires after 0.75–1.25 s. The exchange does not seize the trunk for a new call until the timer has expired. This gives incoming exchange B the time to release the trunk at its end.

Double Seizures. On bothway trunks, double seizures of the trunk by both exchanges can occur. After exchange A seizes a trunk (Fig. 4.2-1), it expects to receive a backward change to off-hook that represents the leading edge of the wink signal. However, this change may also mean that exchange B (at the distant end of the trunk) is sending a seizure signal. After sending a seizure signal, the outgoing exchange thus has to time the duration of the received off-hook. The nominal length of the wink is 140–290 ms. Therefore, the exchanges are arranged to recognize a return on-hook within, say, 100–1000 ms as wink. If the off-hook duration exceeds 1 s, a double seizure has been detected.

There are several ways to deal with a double seizure. For example, both exchanges can be programmed to release the trunk and make a second attempt to set up their calls, trying to seize a trunk in the same trunk group or a trunk in a later-choice group.

4.2.2 Supervision Signaling on FDM Analog Trunks

FDM analog trunks can transfer frequencies between 300 and 3400 Hz. The exchanges indicate the states of the trunk with a 2600-Hz signaling tone. The tone is *in band* (audible) and should be off when the trunk is carrying a call. Therefore, *tone-on-idle* signaling is used:

Trunk State	Tone
On-hook (idle)	Tone on
Off-hook (in use)	Tone off

Sending and Receiving the Signaling Tone. Figure 4.2-2 is a simplified presentation of the signaling circuitry in the four-wire trunk circuits TC4 of a trunk. At each exchange, a *signaling tone source* (STS) supplies all trunk circuits with the 2600-Hz signaling tone. The sending of the tone is controlled by the exchange processor, which controls switch X in the TC4. In the figure, the processor at exchange A has sent an off-hook command to TC4-1, and no tone is sent on the send channel (S) of the trunk. At exchange B, the processor has sent an on-hook command to

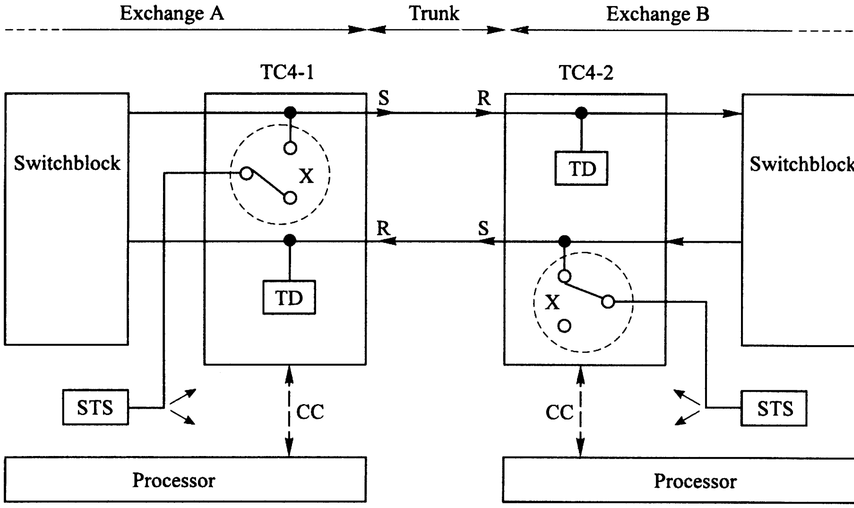


Figure 4.2-2. Four-wire analog trunk circuits (TC4).

TC4-2, and switch X connects the tone on the S channel of the trunk. In the TC4s, a 2600-Hz tone detector (TD) is bridged across the receive channels (R). In this example, the TD in TC4-2 detects no tone and reports to its processor that the trunk is off-hook at exchange A. The TD in TC4-1 detects the tone and reports that the trunk is on-hook at exchange B.

There are several aspects of in-band supervision signaling that pose additional requirements on the circuits of the TC4s.

Blocking Received Signaling Tone. During the conversation, both exchanges indicate off-hook, and no signaling tone is present in either direction. However, in other call states, the signaling tone and other voiceband signals can be present simultaneously on a trunk. Consider a call from subscriber S_1, served by local exchange A, to a subscriber served by local exchange C. The connection passes through intermediate exchange B. Figure 4.2-3(a) shows the transmission path in direction C $\rightarrow$ A only. The called party has not yet answered, and exchange C has attached tone/announcement circuit (TAC) to trunk T_2. The circuit is sending ringing-tone. At points (p), only ringing-tone is present. However, trunk T_2 is on-hook at exchange C, and trunk T_1 is on-hook at exchange B, and at points (q) the ringing-tone and signaling tone are both present.

There are two reasons why TC4-1 and TC4-3 should pass the ringing-tone (or other voiceband signals) but block the signaling tone. In the first place, S_1 should hear ringing-tone only. In the second place, supervision signaling is link-by-link. This means that the signaling tone from B to A on trunk T_1 should be controlled by exchange B. Therefore, TC4-3 has to block the received signaling tone, which otherwise would "leak" into T_1.

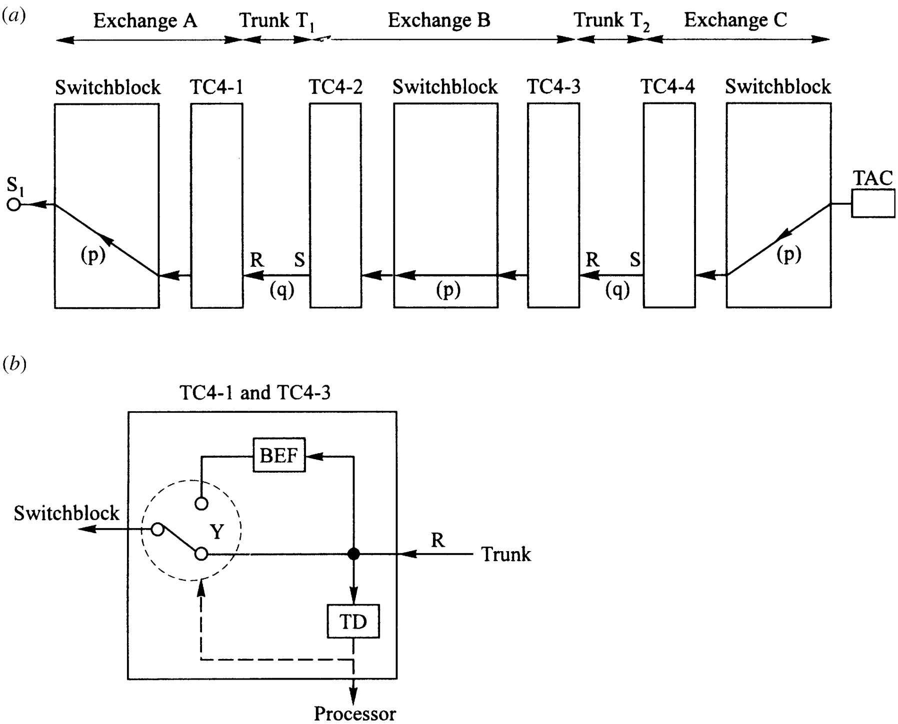


Figure 4.2-3. Blocking received signaling tone: (*a*) transmission path in the direction C → A and (*b*) band elimination filter (BEF).

Figure 4.2-3(*b*) shows how received signaling tone is blocked. *Band elimination filter* (BEF) blocks 2600 Hz but passes other voiceband frequencies. The filter is inserted in receive channel R by switch Y, which is controlled by TD. When TD is not receiving signaling tone, switch Y is in the position shown, and BEF is not in the receive channel. However, when TD receives signaling tone, it sets Y in the other position, and BEF is inserted into the channel.

Protection Against Talk-off. Figure 4.2-4(*a*) shows the transmission path in direction A → C for the connection mentioned earlier. During conversation, the signaling tone is absent at points (p). Speech—or other subscriber communications—can include components around 2600 Hz, which could simulate the signaling tone. In early in-band signaling, a speech fragment from calling subscriber S_1 would occasionally be interpreted by the tone detectors in TC4-2 or TC4-4 as a forward tone-on (on-hook), which is a clear-forward signal. Exchange B or C would then clear the connection. This was known as *talk-off.*

In-band supervision signaling is protected against signal simulation by a combination of two techniques.

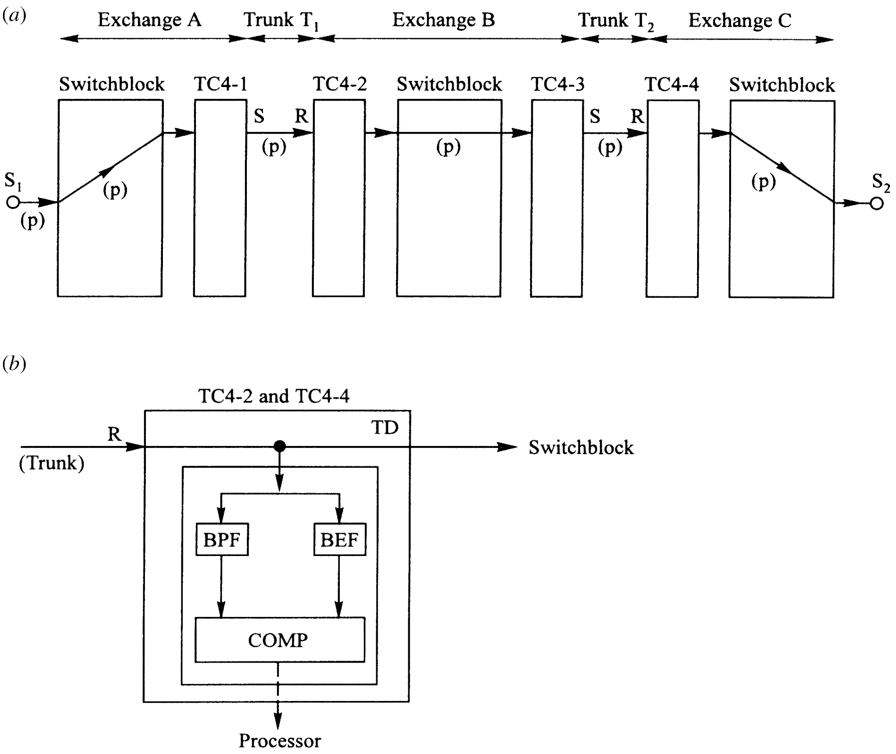


Figure 4.2-4. Protection against talkoff: (*a*) forward transmission path during conversation and (*b*) tone detector (TD).

In the first place, a tone detector (TD) does not simply detect the presence of the 2600-Hz tone. As shown in Fig. 4.2-4(*b*), the signal received from the trunk is fed into band-pass filter BPF that passes a narrow band of frequencies around 2600 Hz only, and to band-elimination filter (BEF) that passes all voiceband frequencies, except the frequencies passed by BPF. Comparator (COMP) compares the signal strengths at the outputs of the filters.

When signaling tone is being received, the output of BPF is stronger than the output of BEF, even when speech is present at the same time. TD thus indicates tone-on (on-hook) to the exchange processor. When only speech is received, the output of BEF is stronger than the output of BPF, because speech power is mainly concentrated in the 300–1000 Hz range, and only a small part of it lies around signaling frequency. In this condition, TD indicates off-hook.

As a second safety measure, exchange processors do not recognize received changes to on-hook or off-hook as valid signals until the new state has persisted during a *recognition time* of a certain length. With the exception of winks, the recognition time for most signals is 30 ms. However, as an extra protection against talk-off, the recognition time for clear-forward signals is 300 ms.

Blue-Box Fraud. In the mid-1960s, the Bell System became aware of "blue-box" fraud on FDM analog trunks [3]. In principle, all signaling systems within-band supervision signaling are vulnerable to this type of fraud, in which a subscriber generates in-band supervision and address signals with a device—called a "blue box"—attached to the subscriber line. The box could generate the 2600-Hz supervision tone and the MF register signals. With the box, a fraudulent subscriber could manipulate the network to avoid charges on long-distance calls.

The Bell System curbed blue-box fraud by vigorous legal prosecution of fraudulent subscribers, and by implementing protective procedures at the exchanges.

4.2.3 Supervision Signaling on Digital (PCM) Trunks

In the example of Section 4.1.3, the supervision signaling for trunks in first-order PCM multiplexes is sent and received by the digital multiplexed ports (DMPs) that attach the PCM multiplexes to the switchblock of the exchange. The DMPs have a control channel to the exchange processor and report the received state (on-hook, off-hook) of the trunks in their multiplexes to the processor. The processor commands the state to be sent out on the trunk.

The DS1 frame format of the North American T1 first-order digital transmission systems (Section 1.5.2) is shown in Fig. 4.2-5. It consists of consecutive frames. Each frame has 24 eight-bit time slots (TSs) and one F bit. Each time slot is associated with a particular trunk.

A series of consecutive frames constitutes a *superframe*, according to the formats described below.

Superframe (SF) Format. Twelve frames form a superframe, in which the frames are numbered from 1 through 12. The DMPs maintain frame and superframe

	Superframe											
Frame #	1	2	3	4	5	6	7	8	9	10	11	12
F bit	1	0	0	0	1	1	0	1	1	1	0	0

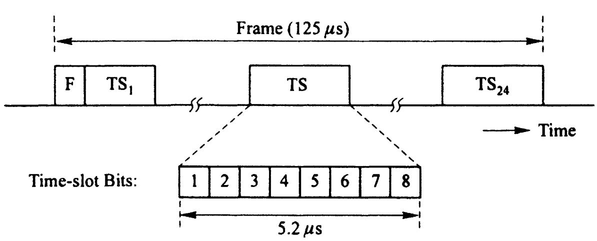


Figure 4.2-5. North American DS1 frame format. Signaling bits are in position 8 of each time slot (TS) during frames 6 and 12.

synchronization with the incoming bit stream by locking onto the F bits of the frames, which exhibit a repeating 12-bit pattern:

	Superframe											
Frame #	1	2	3	4	5	6	7	8	9	10	11	12
F bit	1	0	0	0	1	1	0	1	1	1	0	0

When locked onto this pattern, a DMP can determine the start of each frame, and for each superframe, in the bit stream. In frames 6 and 12, bit 8 (the least significant bits of eight-bit PCM codes) in each of the 24 channels is used for supervision signaling. This is known as *bit robbing*. The effect of bit robbing on the quality of PCM-coded speech is negligible.

The signaling bits in frames 6 and 12 are known as the S_a and S_b bits. The DMPs update their outgoing signaling bits every 1.5 ms (once per superframe), which is sufficiently fast for supervision signaling. The combination of a S_a and S_b bit could indicate four trunk states. However, the S_b bit in each time slot is set equal to the previous S_a bit, resulting in two-state continuous supervision signaling. The bit values 0 and 1 represent, respectively, on-hook and off-hook.

The signaling bits cannot be heard by the subscriber, and the subscriber's speech, or a blue box, cannot corrupt the supervision signals. This avoids the problems associated with in-band signaling.

Extended Superframe (ESF) Format. This is a newer superframe format in which 24 frames, numbered 1 through 24, form a superframe. The F bits of the superframe are used to create three data flows: the framing pattern, a cyclic redundancy check for error detection, and a data link that can be used for operations messages between the entities at each end of the digital facility:

	Extended Superframe																							
Frame #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
F bit	a	b	a	c	a	b	a	c	a	b	a	c	a	b	a	c	a	b	a	c	a	b	a	c

where a = DL (ESF data link), b = CRC (cyclic redundancy check), and c = FPS (frame pattern synchronization). ESF uses six F bits (shown above as "c") for frame and superframe synchronization. The framing pattern is 0 0 1 0 1 1. The six "6" bits transmit a six-bit character that is the result of the CRC calculation on the payload of the previous frame.

The least significant bits of frames 6, 12, 18, and 24 (called A, B, C, D, respectively) of each superframe are "robbed" bits used for signaling, similar to what is done in the SF format.

TABLE 4.2-1 Bell System Multifrequency Address Signals

Signal	Frequencies (Hz)
Digit 1	700 and 900
Digit 2	700 and 1100
Digit 3	900 and 1100
Digit 4	700 and 1300
Digit 5	900 and 1300
Digit 6	1100 and 1300
Digit 7	700 and 1500
Digit 8	900 and 1500
Digit 9	1100 and 1500
Digit 0	1300 and 1500
KP	1100 and 1700
ST	1500 and 1700

4.2.4 Address Signaling

The MF address signals are combinations of two voiceband frequencies—chosen from a set of six frequencies [1,2,5]. The frequency assignments are shown in Table 4.2-1. Only 12 of the 15 possible two-out-of-six codes are used in Signaling System R1. Bell MF signaling uses some of the remaining codes in calls that are set up with operator assistance [5].

Address signaling sequences start with a KP (start-of-pulsing) signal and end with an ST (end-of-pulsing) signal. Signaling sequences received without KP or ST are considered to be mutilated and are discarded by the incoming exchange.

The KP signal has a duration of 90–110 ms. The duration of the other signals is 61–75 ms. The originating exchange sends the complete called number (*en-bloc* address signaling), with silent intervals between signals of 61–75 ms. The intermediate exchanges can use overlap address signaling.

Address Signaling Sequences. In its simplest form, an address signaling sequence conveys the called number only. The called number can be a subscriber number or a national number. Let AC(3), EC(3), and LN(4) represent a three-digit area code, a three-digit exchange code, and a four-digit line number (Section 1.3.1). The address signaling sequences are then KP-EC(3)-LN(4)-ST (subscriber number) or KP-AC(3)-EC(3)-LN(4)-ST (national number).

A calling subscriber dials subscriber numbers for calls inside the numbering plan area (NPA) and dials national numbers for calls outside the NPA. An originating exchange that receives a national called number from the calling subscriber, or an intermediate exchange that receives a national called number from the preceding exchange in the connection, sends out the called number as either a national number or a subscriber number. This depends on whether the exchange has seized an outgoing trunk to an exchange outside or inside the NPA of the called party.

Automatic Number Identification. In the early years of subscriber-dialed long-distance calling, most local exchanges were not equipped to produce billing records for these calls, and the billing records were generated by the first intermediate (toll) exchange in the connection. The local exchange would send both the called and calling numbers, and the toll exchange would handle the charging for the call. After the end of the call, the toll exchange would generate a billing record that included both numbers, the date of the call, and the times of answer and call clearing. The sending of calling numbers is known as *automatic number identification* (ANI).

After the breakup of the Bell System, calls between subscribers in different local (LATA) networks are billed by either the LATA of the calling party or the interexchange carrier (IC). This is a matter of mutual agreement between the LATA and IC.

Figure 4.2-6 shows the supervision and address signals for an inter-LATA call originated by subscriber S. The call is to be billed by IC exchange B.

A subscriber can designate a "default" IC for his inter-LATA calls. This information is stored at the local exchange. If a subscriber just dials a called number, the exchange routes the call to an exchange of his default IC. If subscriber S desires a different carrier for the call, he dials a prefix 10*XXX*, where *XXX* identifies the IC.

After the subscriber has dialed

1-AC(3)-EC(3)	(called national number, default IC), or
101*XXXX*-1-AC(3)-EC(3)	(called national number, specified IC), or
EC(3)	(called subscriber number, default IC), or
$101XXXX - EC(3)$	(called subscriber number, specified IC)

exchange A knows the desired IC and the nearest exchange (B) of that IC. This example assumes that a direct trunk group connects exchanges A and B. Exchange A seizes a trunk T in this group and sends a seizure signal. After receiving

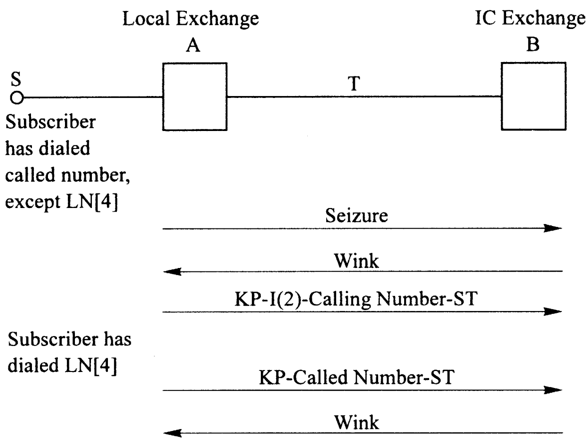


Figure 4.2-6. Transfer of calling and called numbers to exchange B of interexchange carrier.

the wink signal, A first sends the national number of the calling subscriber:

$$\text{KP-I(2)-AC}'\text{(3)-EC}'\text{(3)-LN}'\text{(4)-ST}$$

When the subscriber has finished dialing LN(4), the last four digits of the called number, exchange A sends a second digit sequence, which identifies the called national or subscriber number: KP-AC(3)-EC(3)-LN(4)-ST or KP-EC(3)-LN(4)-ST. Exchange B then acknowledges the receipt of both numbers with a wink signal.

Sending the calling number first minimizes the elapsed time from the end of address signaling by the calling subscriber to the end of address signaling by the local exchange.

The codes in the two information digits I(2) ahead of the calling number characterize the calling line:

I(2) = 00	Identified subscriber line
I(2) = 02	ANI failure (calling number not included)
I(2) = 06	Call from hotel without room dentification
I(2) = 10	Test call

When I(2) = 02 or 06, an operator at exchange B verbally obtains the calling number.

4.2.5 Failed Setups

Bell MF signaling does not include backward signals to indicate that the setup of a connection has failed. The exchange where the failure occurs sends a tone (busy-tone, reorder-tone) or announcement, and the calling party disconnects (sends a clear-forward to the originating exchange). The originating exchange then initiates the release of the connection.

4.3 SIGNALING SYSTEM NO. 5

This signaling system was deve1oped jointly by the U.K. post office and Bell Laboratories and is similar to Bell System MF signaling. It was adopted in 1964 by CCITT for use in the international network, and for that reason it was initially called CCITT No. 5 Signaling.

Signaling System No. 5 has been designed especially to operate on TASI-equipped analog trunks (Section 1.6.4). Supervision and address signaling are both link-by-link and in-band.

Once used extensively on long international trunks (underwater transoceanic trunks, satellite trunks), Signaling System No. 5 is not widely used anymore, so we refer the reader to the relevant ITU-T recommendations [6] for the details of the protocol.

4.4 MFC-R2 SIGNALING

R2 signaling, also known as Multifrequency Compelled (MFC), was developed co-operatively by European telecommunication equipment manufacturers and the European Telecommunication Standards Institute (ETSI). It was introduced in the 1960s and is still used in many national networks in Europe, Latin America, Australia, and Asia.

A few years after the introduction of MFC, CCITT (now ITU-T) defined a version for use in the international network. This international version is officially called Signaling System R2. Both the international MFC system and the national MFC system (which exists in several country-specific versions) are commonly referred to as MFC-R2 or R2 signaling systems [1,4,7].

R2 signaling can be used on two-wire analog trunks and on four-wire analog and digital trunks. It cannot be used on TASI-equipped trunks or on trunks carried by satellite transmission systems. This limits the application of R2 to relatively short international trunks (Sections 4.4.1 and 4.4.3).

Compared with Bell System MF signaling and Signaling System No. 5, the most important difference of R2 is its *register* (address) signaling.

4.4.1 Supervision Signaling on FDM Analog Trunks

This section describes the supervision signaling of national and international R2, for four-wire FDM analog trunks [4].

The signaling is *out-of-band*: the bandwidth of the channels in FDM transmission systems is divided into a 300–3400 Hz band for the subscriber's speech (or other communications) and a narrow band centered at the signaling frequency $f = 3825$ Hz. This separates signaling tone and speech, and avoids the problems associated with in-band supervision signaling (4.2.2).

Trunk Circuit. Figure 4.4-1 shows a four-wire trunk circuit TC4 for out-of-band supervision signaling. *Low-pass filters* (LPFs) that block frequencies above 3400 Hz separate speech from signaling tone, and vice versa.

Switch X inserts the signaling tone into the send (S_2) channel on command from the exchange processor, and tone detector (TD) reports the presence or absence of the signaling tone on the receive channel R_2 to the processor.

International R2 Supervision Signals. The supervision signaling of Signaling System R2, intended for one-way analog trunks only, is two-state, tone-on-idle (as in Bell System MF signaling). For this reason, R2 signaling cannot be used on

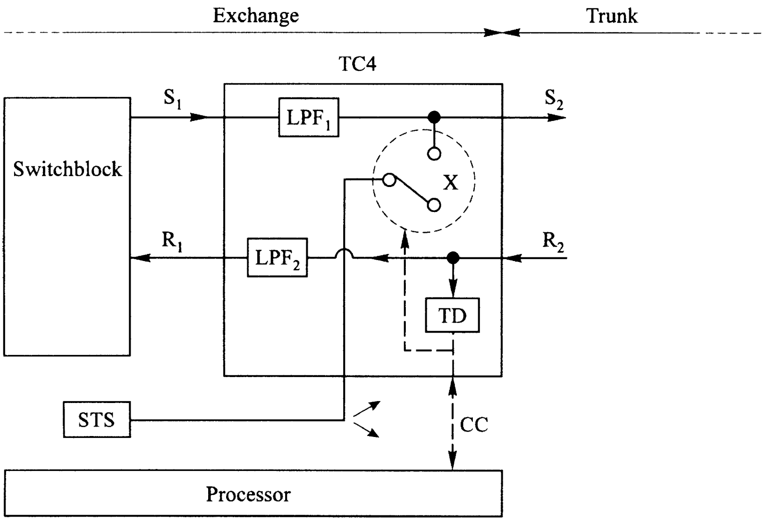


Figure 4.4-1. Four-Wire analog trunk circuit (TC4) for out-of-band supervision signaling.

TASI-equipped trunks. Nominal signal recognition times are 40 ms. The signals and the corresponding changes in trunk state are:

Forward Signals	State Change
Seizure	Tone-on to tone-off
Clear-forward	Tone-off to tone-on

Backward Signals	State Change
Answer	Tone-on to tone-off
Clear-back	Tone-off to tone-on
Release-guard	450-ms Tone-off pulse, or tone-off to tone-on
Blocking	Tone-on to tone-off

The system does not include a proceed-to-send signal. The blocking signal is a backward transition to tone-off when the trunk has not been seized by the outgoing exchange. It is sent by the incoming exchange and requests the outgoing exchange to suspend seizing the trunk for new calls, because the incoming exchange is performing maintenance on the trunk.

The release-guard signal is sent by the incoming exchange after it has received a clear-forward signal. It indicates that the incoming exchange has released the trunk at its end, and that the outgoing exchange can therefore seize the trunk for a new call.

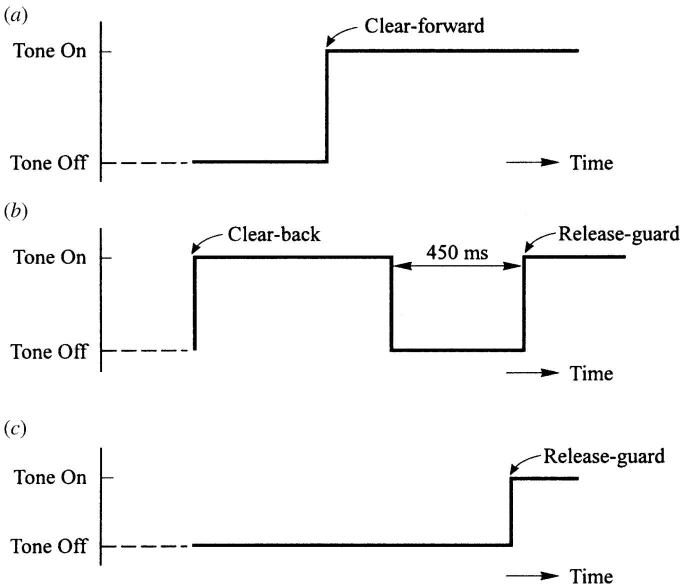


Figure 4.4-2. Release-guard signals: (*a*) clear-forward, (*b*) release-guard signal when clear-back signal has been sent, and (*c*) release-guard signal when clear-back signal has not been sent.

As shown in Fig. 4.4-2, if the incoming exchange receives a clear-forward after it has sent a clear-back, the release-guard signal is a 450-ms tone-off pulse (b). If the incoming exchange has not sent a clear-back when it receives the clear-forward, the release-guard signal is a change to on-hook, 450 ms after the receipt of the clear-forward (c).

National R2 Supervision Signals. Most national R2 systems use pulsed out-of-band supervision signals. There are several country-specific pulsed signaling systems, for example:

Forward Signals	**Pulse Duration (ms)**
Seizure	150
Clear-forward	600

Backward Signals	**Pulse Duration (ms)**
Answer	150
Clear-back	600
Release-guard	600
Blocking	Continuous

4.4.2 Supervision Signaling on Digital Trunks

Networks that use national or international R2 signaling use El first-order transmission systems for digital trunks [4]. The bit streams on these multiplexes are organized in frames that are transmitted at a rate of 8000 frames/s. Each frame has 32 eight-bit time slots, numbered from TS_0 through TS_{31}—Fig. 1.5-4(*b*). Time slots TS_1 through TS_{15} and TS_{17} through TS_{31} carry PCM-encoded speech, or 64-kb/s subscriber data, for 30 trunks. For frame alignment, bits 2 through 8 of TS_0 have the fixed pattern 0011011.

A superframe consists of 16 consecutive frames, numbered from 0 through 15. For superframe alignment, bits 1 through 4 in TS_{16} of frame 0 are coded 0000. TS_{16} in frames 1 through 15 carries four status bits (a,b,c,d) for the trunks:

< Bits in Time Slot 16 >

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
	a	**b**	**c**	**d**	**a**	**b**	**c**	**d**
Frame 1	<Trunk 1>				<Trunk 17>			
Frame 2	<Trunk 2>				<Trunk 18>			
Frame 3	<Trunk 3>				<Trunk 19>			
⋮	⋮				⋮			
Frame 15	<Trunk 15>				<Trunk 31>			

The supervision signaling for digital international R2 trunks is continuous, with two forward and three backward trunk states that are represented by bits a_f, b_f, and a_b, b_b, respectively. Bits c and d are not used and are set to 0 and 1.

The signaling can be applied to one-way and bothway trunks. On one-way trunks, only one exchange can seize a trunk and send forward bits a_f and b_f, and the other exchange sends backward bits a_b and b_b. On bothway trunks, the roles of the exchanges vary from call to call, depending on which exchange seizes the trunk.

The idle state at both ends of a trunk is represented by a,b = 1,0. The supervision signals are represented by changes in bit patterns:

Forward Signals	**Change**
Seizure	a_f, b_f: 1,0 → 0,0
Clear-forward	a_f, b_f: 0,0 → 1,0

Backward Signals	**Change**
Seizure acknowledgment	a_b, b_b: 1,0 → 1,1
Answer	a_b, b_b: 1,1 → 0,1
Clear-back	a_b, b_b: 0,1 → 1,1
Release-guard	a_b, b_b: 1,1 → 1,0
	or a_b, b_b: 0,1 → 1,0

Blocking. An exchange can block an idle trunk by changing its status bits from a,b = 1,0 to a,b = 1,1. Exchanges do not seize trunks that are in this state at the distant end. To end blocking, the exchange returns the bits to a,b = 1,0 (idle).

Double Seizure. After sending a seizure signal, the outgoing exchange expects to receive a change to $a_b,b_b = 1,1$ from the incoming exchange. The response $a_b,b_b = 0,0$ indicates a double seizure. Both exchanges then abort their call setups and—depending on the procedures of the telecom—either make a second attempt or abort the call setups, sending congestion indications to the calling subscribers.

4.4.3 Interregister Signaling

In R2 signaling, the equipment units at the exchanges that send and receive digits, and the signaling between these units, are usually referred to as *registers*, and *inter-register signaling.*

R2 uses forward and backward in-band MF (multifrequency) signals. On digital trunks, these signals are PCM-coded. Signaling is *compelled*: a forward MF signal, sent by an *outgoing register* (OR) at outgoing exchange A, is held "on" until the receipt of a backward MF acknowledgment from *incoming register* (IR) at incoming exchange B—see Fig. 4.4-3. R2 registers are transceivers: they both send and receive register signals.

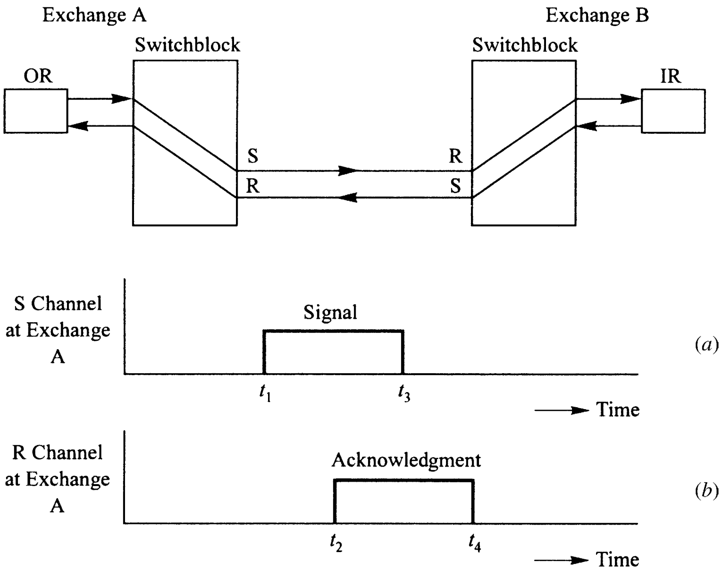


Figure 4.4-3. Compelled interregister signal and acknowledgment.

Compelled interregister signaling is not practical for trunks carried on satellite transmission systems, because of their long propagation times (about 600 ms). The time required for the transmission of one digit and its acknowledgment would be on the order of 1.2 s, and this would result in extremely slow signaling.

Interregister Signals. The forward and backward signals consist of two voice-band frequencies, selected from a set of six—see Table 4.4-1. The forward and backward frequency sets are different. This is necessary to make the register signaling suitable for two-wire analog trunks.

Because of the two frequency sets, two register types are needed. In Fig. 4.4-3, outgoing register (OR) sends forward frequencies and receives backward frequencies. The reverse holds for incoming register (IR).

Groups of Interregister Signals. A particular forward or backward signal can have several meanings. We speak of group I, group II, and, in some countries, group III *forward* signals, and of group A and group B *backward* signals. A signal is denoted by its "group" and its "value." For example, we denote a group II signal with value 7 by II-7.

The group I forward signals represent the digits of the called party number. Group II signals indicate the category of the calling party, and group III signals, which are used in some national networks only, represent the digits of the calling party's number.

In MFC-R2 signaling, the *incoming exchange* controls the signaling sequence. A group A signal requests a particular next forward signal or indicates that register signaling has ended. Group A signals can be sent by intermediate exchanges and by the terminating local exchange.

TABLE 4.4-1 National and International R2 Interregister Signal Frequencies

Signal Value	Forward (Hz)	Backward (Hz)
1	1380 and 1500	1140 and 1020
2	1380 and 1620	1140 and 900
3	1500 and 1620	1020 and 900
4	1380 and 1740	1140 and 780
5	1500 and 1740	1020 and 780
6	1620 and 1740	900 and 780
7	1380 and 1860	1140 and 660
8	1500 and 1860	1020 and 660
9	1620 and 1860	900 and 660
10	1740 and 1860	780 and 660
11	1380 and 1980	1140 and 540
12	1500 and 1980	1020 and 540
13	1620 and 1980	900 and 540
14	1740 and 1980	780 and 540
15	1860 and 1980	660 and 540

Source: Rec. Q. 441. Courtesy of ITU-T.

The group B backward signals are sent by the terminating local exchange only. They acknowledge a forward signal and convey call-charging instructions and called-party status.

The incoming and outgoing exchanges must know the type of the signal that is being received. To accomplish this, a R2 register signaling sequence follows certain rules:

- The first signal received by an incoming exchange is a group I signal.
- The outgoing exchange interprets received backward signals as group A signals, until it receives a group A signal that indicates that the next backward signal will be a group B signal. The receipt of a group B signal always ends the signal sequence.

4.4.4 National R2 Interregister Signaling Sequences

We now explore a few register signaling sequences in a national network. In these examples, the meanings of the forward and backward register signals, which vary somewhat from country to country, are as listed in Tables 4.4-2 and 4.4-3.

Figure 4.4-4 shows the signaling on trunk T, for a call from subscriber S_1 to S_2, whose subscriber number is 34-5678. Exchange X has received the called number from S_1. It seizes trunk T and sends a seizure signal. The exchange also connects an outgoing register (OR) to the trunk and orders it to send I-3 (the first digit of the called number). When exchange Z receives the seizure signal, it connects an incoming register (IR) to the trunk. The register receives the I-3 and acknowledges with A-1 (send next digit). The acknowledgment of the first digit indicates to exchange X that an incoming register has been connected to the trunk (this is why R2 signaling does not include a proceed-to-send signal). Exchange X sends the subsequent digits, and exchange Z acknowledges the second through fifth digits with A-1. On receipt of the sixth digit (I-8), exchange Z knows that the called number is complete and acknowledges with A-3, which requests the calling party category, and indicates that the next backward signal is a group B signal. After receiving calling category (II-2), exchange Z sends a group B signal that contains information on the called party status and on charging. The group B signal ends the interregister signaling.

End-to-End Interregister Signaling. Now consider the signaling for a call from S_1 to S_2 that is routed via intermediate exchange Y—see Fig. 4.4-5. The subscriber number of S_2 is again 34-5678. In some countries, R2 register signaling is link-by-link. This means that exchange Y receives the entire called number from exchange X, then seizes trunk T_2, and sends the number to exchange Z.

However, in most national networks, R2 register signaling is end-to-end. In this mode, outgoing register (OR) in originating exchange X communicates successively with incoming registers (IRs) in exchanges Y and Z. The initial register signaling is

TABLE 4.4-2 Example of National R2 Forward Interregister Signals

Group I: Digits in the Called Number
I-1 Digit 1
I-2 Digit 2
I-3 Digit 3
I-4 Digit 4
I-5 Digit 5
I-6 Digit 6
I-7 Digit 7
I-8 Digit 8
I-9 Digit 9
I-10 Digit 0
I-15 End of called number
I-11 through I-14 are not used
Group II: Calling Party Category
II-1 Operator with trunk-offering
II-2 Subscriber
II-3 Pay-phone
II-4 through II-15 are not used
Group III: Digits in the Calling Number
III-1 Digit 1
III-2 Digit 2
III-3 Digit 3
III-4 Digit 4
III-5 Digit 5
III-6 Digit 6
III-7 Digit 7
III-8 Digit 8
III-9 Digit 9
III-10 Digit 0
III-15 End of called number
III-11 through III-14 are not used

between exchanges X and Y. Having received the called number 34-5678 from S_1, exchange X seizes trunk T_1 and starts sending the called number. After receiving the exchange code EC = 34, exchange Y knows that Z is the terminating exchange for the call. It therefore acknowledges the I-4 with A-2, which is a request to restart sending the called number.

Exchange Y then disconnects its IR from T_1, seizes trunk T_2, sends a seizure signal on T_2, and cuts through a path between the trunks. From this point on, the register signaling is between exchanges X and Z. Exchange X responds to the A-2 by sending I-3 (the first digit of the called number). Exchange Z connects an IR to T_2 and, when the register receives the I-3, it requests the next digit by acknowledging

TABLE 4.4-3 Example of R2 National Backward Interregister Signals

	Group A
A-1	Send next digit of called number
A-2	Resend first digit of called number
A-3	Send calling line category and prepare to receive a group B signal
A-4	Congestion
A-5	Send calling line category
A-6	Not used
A-7	Send next digit of calling number
A-8	Digit *n* of called number received; resend digit ($n - 1$)
A-9	Digit *n* of called number received; resend digit ($n - 2$)
A-10	through A-15 are not used
	Group B
B-1	Called subscriber idle, charge
B-2	Called subscriber busy
B-5	Called subscriber idle, do not charge
B-6	Called party idle, call to be held under control of called subscriber
B-7	Vacant number received
B-3, B-4, and B-8 through B-15 are not used	

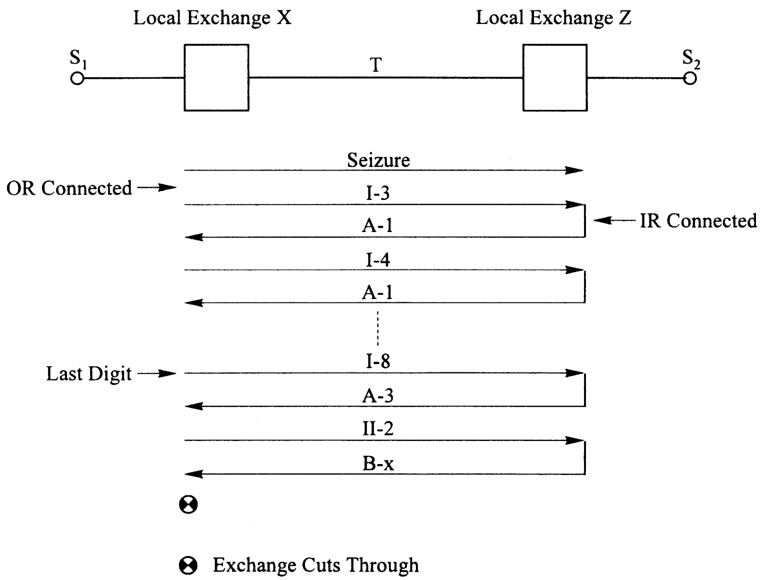


Figure 4.4-4. R2 interregister signaling example.

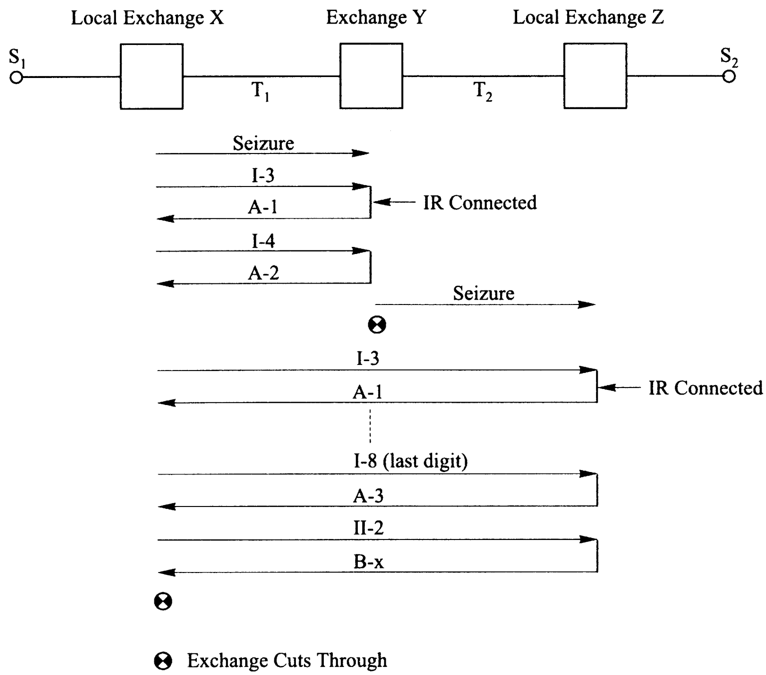


Figure 4.4-5. End-to-end interregister signaling.

with A-1. The second through fifth digits are acknowledged in the same way, and the sixth (final) digit is acknowledged with A-3. Exchange X then sends the calling party category (II-2), and exchange Z ends register signaling with a group B signal.

R2 end-to-end signaling in national networks is also possible on connections that pass through several intermediate exchanges.

4.4.5 R2 Supported Features in National Networks

With end-to-end R2 interregister signaling, information about the status and nature of the called party can be sent from the terminating to the originating local exchange. In addition, the calling party category (subscriber, operator, etc.) is sent from the originating to the terminating exchange. The combination of these procedures supports a number of national network features that cannot be provided by Bell System MF signaling. Some of these features are outlined below, using the call of Fig. 4.4-5 as an example.

Free Calls. Certain lines on a local exchange may be marked as "free" destinations. During the register signaling on a call to a free destination at exchange Z, this exchange can indicate to originating exchange X that the call should not be charged (signal B-5).

Called Party Hold. Specific lines on a local exchange can be marked as having "called party holding." On calls to such lines, terminating exchange Z ends the signaling with B-6. This requests originating exchange X to hold the connection until it receives a clear-back (line signal). Lines of police and fire departments are often marked as lines with called party holding.

Malicious Call Tracing. Lines on a local exchange can be marked as "subjected to malicious calls." On calls to such lines, terminating exchange Z requests the digits of the calling party number from exchange X, with a series of A-7 signals, and stores the number. If the call is malicious, the called party alerts the local exchange, usually by a hookswitch flash—see Section 3.3.2. In response, exchange Z prints a record that includes the date and time of day of the call, and the numbers of the calling and called parties.

Trunk Offering. This allows an operator to break in on a busy line. Suppose that a subscriber, say, S_1, calls a subscriber S_2 on exchange Z and that this subscriber is busy. In an emergency situation, S_1 can call a "trunk offering" operator for assistance. The operator then places a call to S_2. Say that this call arrives at exchange Z on a trunk T. Exchange Z receives the calling party category signal II-1 and recognizes that the calling party is a "trunk-offering" operator. It then bridges trunk T onto the existing call of subscriber S_2. The operator then informs the subscriber about the emergency call.

Release of Connection When Setup Fails. When intermediate exchange Y or terminating exchange Z cannot extend the call setup, it sends a backward register signal that indicates the nature of the problem. For example, intermediate exchange Y can indicate that no trunk to Z is available (A-4), and exchange Z can indicate that the called line is busy (B-2), or that the called number is not in use (B-7). In response, originating exchange X initiates the release of the connection and connects the calling party to a suitable tone or announcement source. This expedites the release of trunks T_1 and T_2.

4.4.6 International R2 Signaling

Signaling System R2 is the international version of R2 signaling, specified by ITU-T [4] for use in the international network on one-way four-wire analog (FDM) trunks, and on one-way and bothway digital (PCM) trunks [4].

Supervision signaling is as described in Sections 4.4.1 and 4.4.2. The supervision signaling on analog trunks is continuous "tone-on-idle." This makes the system unsuitable for TASI-equipped trunks (Section 1.6.4) because, when a trunk is idle, the tone-on condition would cause the occupation of a pair of bearer channels.

Signaling System R2 is also used very rarely on satellite trunks, because each compelled MF interregister signal would involve two one-way propagation delays of about 600 ms each, which would result in very slow register signaling.

4.4.7 International R2 Interregister Signaling

The signaling is an adaptation of R2 national signaling to the requirements of the international network and includes a number of specific international signals.

The forward register signals are divided into three groups—see Table 4.4-4.

Group 0 Signals. The first signal received by an incoming exchange is a group 0 signal. These signals differentiate calls that terminate in-country from transit calls. The first signal in a *terminal* seizure is 0-1 through 0-5, or 0-10 (which represent Z-digit values), or 0-13 (indicating a call from test equipment at the outgoing ISC).

TABLE 4.4-4 International R2 Forward Interregister Signals

Group 0	
0-1	French-speaking operator
0-2	English-speaking operator
0-3	German-speaking operator
0-4	Russian-speaking operator
0-5	Spanish-speaking operator
0-10	Subscriber
0-11	Country code indicator; outgoing half-echo suppressor required
0-12	Country code indicator; no echo suppressor required
0-13	Call by automatic test equipment
0-14	Country code indicator; outgoing half-echo suppressor included
0-6 through 0-9, and 0-15 are not used	
Group I: Digits in Called Number	
I-1	Digit 1
I-2	Digit 2
I-3	Digit 3
I-4	Digit 4
I-5	Digit 5
I-6	Digit 6
I-7	Digit 7
I-8	Digit 8
I-9	Digit 9
I-10	Digit 0
I-11	Code 11
I-12	Code 12
I-13	Address code for test equipment at incoming exchange
I-14	Incoming half-echo suppressor required
I-15	End of called number
Group II: Calling Party Category	
II-7	Subscriber, or operator without forward transfer
II-8	Data transmission
II-9	Subscriber with priority
II-1	through II-6, and II-10 through II-I5 are not used

Source: Rec. Q. 441. Courtesy of ITU-T.

TABLE 4.4-5 International R2 Backward Interregister Signals

	Group A
A-1	Send next digit $(n+1)$
A-2	Resend digit $(n-1)$
A-3	Address complete, prepare to receive a group B signal
A-4	Congestion in national network
A-5	Send calling patty category
A-6	Address complete, end of register signaling
A-7	Resend digit $(n-2)$
A-8	Resend digit $(n-3)$
A-11	Sent country code indicator (transit seizure)
A-12	Send Z-digit (terminal seizure)
A-13	Send nature of circuit
A-14	Send echo suppressor information
A-15	Congestion at International Switching Center, or all trunks busy
A-9	and A-10 are not used
	Group B
B-2	Send special information tone
B-3	Subscriber line busy
B-4	Congestion
B-5	Unallocated number
B-6	Subscriber idle, charge
B-7	Subscriber line idle, do not charge
B-8	Subscriber line out of service
B-9	through B-15 are not used

Source: Rec. Q. 441. Courtesy of ITU-T.

Signals 0-11, 0-12, and 0-14 indicate a *transit* seizure. These signals imply that the called number includes a country code and are also known as *country code indicators*. In addition, the signals contain information for the control of echo suppressors.

Group I Signals. The forward signals I-1 through I-10 represent the digits of the called number. I-11 and I-12 are addresses for groups of international operators, and I-13 is the first digit of addresses that specify a particular type of test equipment at the ISC.

Group II Signals. These forward signals indicate the calling party category and are sent to exchanges in the destination country when end-to-end signaling is possible. In practice, only II-7 is used.

The backward signals are listed in Table 4.4-5.

Group A Signals. Signal A-1 requests the next address digit. Signals A-2, A-7, A-8, A-11, and A-12 are sent, just before cut-through, by a transit ISC that has selected a R2 outgoing trunk. They indicate the next signal to be sent by the originating ISC.

A-3 and A-6 indicate that the complete called number has been received. A-6 ends the register signaling, and A-3 requests the calling party category, which will be acknowledged by a group B signal.

A-4 and A-15 indicate that the call cannot be set up and are requests to the outgoing ISC to release the connection.

Group B Signals. These signals convey information on the nature and status of the called subscriber. These signals can be sent only by an ISC that receives this information from a terminating local exchange in its national network.

4.4.8 International R2 Interregister Signaling Example

We consider the international call shown in Fig. 4.4-6. The call is routed via transit country B. Trunks T_1 and T_2 have R2 signaling. Even if the national networks in countries A and C use R2 signaling, there are three signaling sections in the connection. This is necessary because local exchanges (e.g., X) are not equipped to handle international signaling procedures, and because some national R2 register signals are not identical to the R2 register signals. If the networks of countries A and C use R2 signaling, end-to-end signaling is possible in signaling section 1 (between local exchange X and ISC-P), in section 2 (between ISC-P and ISC-R), and in section 3 (between ISC-R and the terminating local exchange Z).

The signaling in the international network is shown in Fig. 4.4-7 (see Tables 4.4-4 and 4.4-5). The called international number is 34-67-412-1093, where 34 is the country code of C, and 67-412-1093 is the national number of S_2.

Outgoing ISC-P seizes trunk T_1 to ISC-Q and starts by sending country code indicator O-12, indicating a "transit" seizure (the call destination is not in country B). ISC-Q acknowledges with A-1, requesting the first digit of the called international number. We assume that ISC-R is the only ISC in country C. Therefore, ISC-Q can make a route decision after receipt of the country code CC (34). In this example, it seizes a direct trunk T_2 to ISC-R. Since this ISC is in the destination country, it should receive a terminal seizure signal, followed by the called national

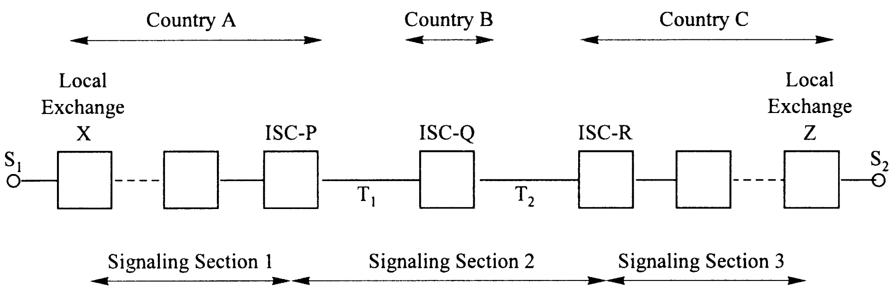


Figure 4.4-6. International connection. International R2 signaling on T_1 and T_2.

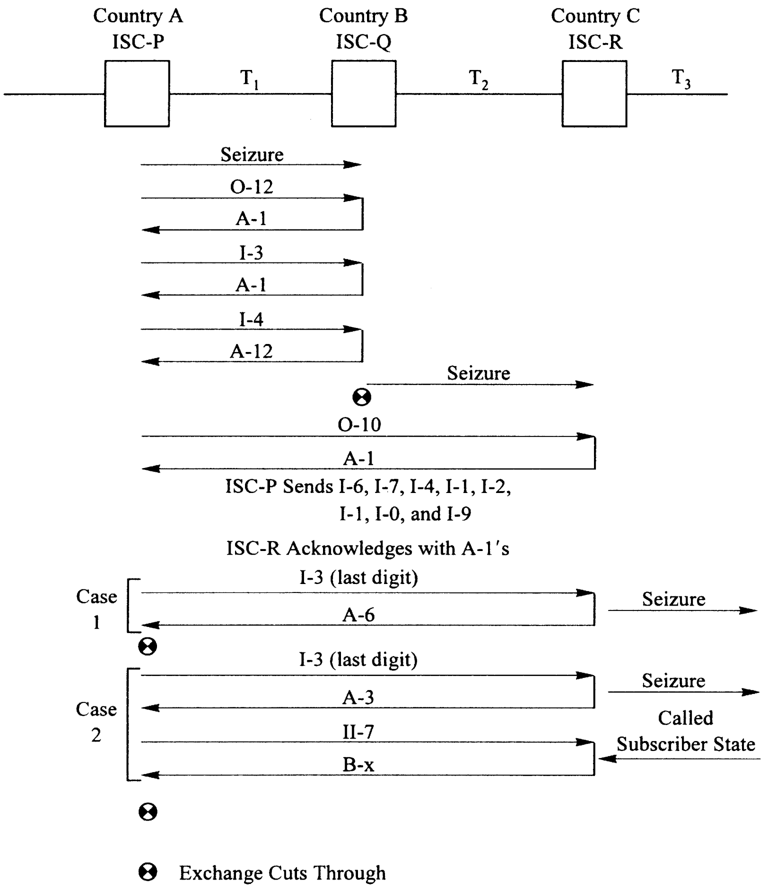


Figure 4.4-7. Signaling in the international section of the connection of Fig. 4.4-6.

number. ISC-Q therefore acknowledges the final digit of the country code with A-12 and then cuts through. ISC-P starts its register signaling with ISC-R by sending O-10 (terminal seizure, call originated by a subscriber) and, prompted by successive A-1 signals, sends the national number. ISC-R knows the lengths of national numbers in its country and acknowledges all received digits except the last one with A-1.

The acknowledgment of the final digit of the called number depends on whether ISC-R can obtain information about the called subscriber from its national network.

Case 1. The national network of the destination country cannot signal the status of the called party to its incoming ISC-R. This was the case in the United States prior to the introduction of common-channel signaling. After receiving the last digit (I-3) of the called number, ISC-R seizes national trunk T_3 and acknowledges the I-3 with A-6. This ends the register signaling in the international network. ISC-P cuts through, and the calling subscriber eventually receives ringing-tone, busy tone, or an announcement from an exchange in country C.

Case 2. ISC-R can obtain the condition of the called subscriber (e.g., when country C uses national R2 with end-to-end register signaling). In this case, ISC-R seizes trunk T_3 after receipt of the last digit of the called number and acknowledges with A-3. This requests the calling party category and informs ISC-P that the next acknowledgment it will receive is a group B signal.

ISC-P then sends II-7 (call originated by a subscriber). When the connection in the destination country has been set up, and ISC-R has received information on the status of the called subscriber, it sends the appropriate group B signal. This ends the register signaling between ISC-P and ISC-R.

4.4.9 Pulsed Group A Signals

Up to this point, a backward group A signals always acknowledges a received forward signal. National and international R2 signaling also use pulsed group A signals. The pulsed signals (pulse duration: 100–200 ms) do not acknowledge a received forward signal and are sent when the incoming exchange has information for the outgoing exchange at a time that the outgoing exchange is not sending a forward signal.

The use of pulsed signals is illustrated with the international call of Fig. 4.4-6. We assume that originating country A has R2 signaling and consider the signaling between originating exchange X and ISC-P. The forward and backward interregister signals of Tables 4.4-2 and 4.4-3 are used.

Originating local exchanges cannot determine whether an international number received from the calling party is complete. Moreover, early ISCs were not equipped with stored information on the lengths of international numbers. We assume that ISC-P is such an exchange. Figure 4.4-8 shows the end-to-end interregister signaling between X and ISC-P, from the time that exchange Y has seized trunk T. Exchange X then starts to send the digits of the international called number (34-67-412-1093) received from S_1. Since ISC-P cannot determine when the received number is complete, it acknowledges each group I signal with an A-l and also starts (or restarts) a 5-s timer on the receipt of each signal. When X receives the A-1 acknowledgment of the eleventh digit (I-3), it falls silent, because it has sent all received digits. The timer at ISC-P then times out, and ISC-P assumes that the called number is complete.

It now sends a pulsed A-3 signal to exchange X, requesting the calling party category. It also seizes an outgoing trunk in the international network. The A-3 also indicates that the next acknowledgment will be a group B signal. ISC-P acknowledges the II-2 with a B-l. This ends the register signaling between X and ISC-P, and originating exchange X cuts through.

The B-1 (called subscriber free, charge) is sent by convention: ISC-P does not know whether it will receive information about the call setup or the status of the called subscriber. If ISC-P ends its register signaling in the international network without obtaining this information or receiving an indication that the called subscriber is free, it cuts through. The calling subscriber then receives an audible signal from an intermediate exchange, or from terminating local exchange Z. If ISC-P

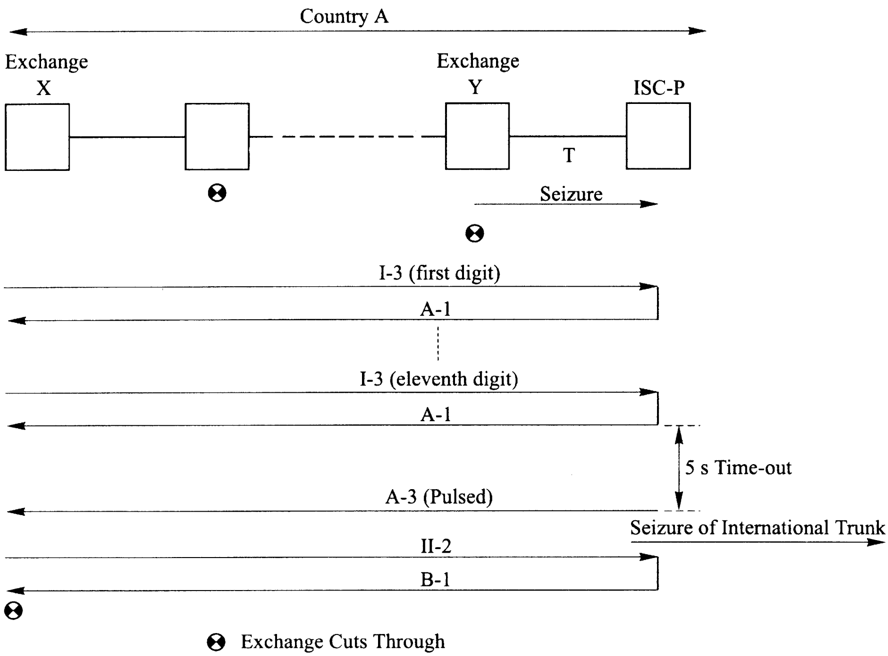


Figure 4.4-8. Pulsed group A signal in Section 1 of Fig. 4.4-6.

receives an indication during its international register signaling that the call cannot be set up (called subscriber busy, no trunks available, etc.), it connects its incoming trunk T to a tone or announcement source.

4.5 ACRONYMS

ac	Alternating current
AC	Area code
ANI	Automatic number identification
BEF	Band-elimination filter
BPF	Band-pass filter
CAS	Channel-associated interexchange signaling
CC	Control channel
CCITT	International Telegraph and Telephone Consultative Committee
COMP	Comparator
DC	Direct current
DDD	Direct distance dialing
DMP	Digital multiplexed port
DS1	Frame format of American first-order digital multiplex

EC	Exchange code
ETSI	European Telecommunications Standards Institute
FDM	Frequency-division multiplexing
IC	Interexchange carrier
IR	Incoming register
ISC	International switching center
ITU	International Telecommunication Union
ITU-T	Telecommunication Standardization Sector of ITU
KP	First signal in MF register signaling sequence
LATA	Local access and transport area
LN	Line number
LPF	Low-pass filter
MF	Multifrequency
MFC	Multifrequency Compelled
NN	National number
OR	Outgoing register
PCM	Pulse code modulation
POTS	Plain old telephone service
R	Receive channel
RCV	Digit receiver
S	Send channel
SPC	Stored program controlled
SN	Subscriber number
SND	Digit sender
ST	Final signal in MF register signaling sequence
STS	Signaling tone source
TAC	Tone and announcement circuit
TASI	Time assignment speech interpolation
TC4	Four-wire analog trunk circuit
TD	Signaling tone detector
TDM	Time-division multiplexing
TS	Time slot

4.6 REFERENCES

1. S. Welch, *Signalling in Telecommunications Networks*, Peter Peregrinus Ltd., Stevenage, UK, 1981.
2. C. Breen and C. A. Dahlbom, Signalling Systems for Control of Telephone Switching, *Bell Syst. Tech. J.*, **39**, 1960.

3. *A History of Engineering and Science in the Bell System*, Bell Telephone Laboratories, 1982.
4. *Specifications of Signalling Systems R1 and R2*, Rec. Q.310–Q.331 (R1 signaling) and Q.400–Q.480 (R2 signaling), ITU-T, Geneva, 1985.
5. *Notes on the BOC Intra-Lata Networks*, AT&T, New York, 1983.
6. *Specifications of Signalling Systems No. 4 and No. 5*, Rec. Q.140–Q.164, ITU-T, Geneva, 1985.
7. M. den Hertog, "Interregister Multifrequency Code Signalling for Telephone Switching in Europe," *Elec. Commun.*, **38**, 1963.

5

INTRODUCTION TO COMMON-CHANNEL SIGNALING

In *channel-associated signaling* (CAS) systems, the signaling information for a trunk is carried by the trunk itself. In *common-channel signaling* (CCS), a common signaling link (SL) carries *signaling messages* for a number of trunks. Just as multifrequency (MF) signaling became feasible with the introduction of the second-generation (common-control) switching systems, CCS was developed for the third-generation (stored-program controlled—SPC) exchanges that were introduced in the 1960s.

There were several reasons for the move from multifrequency signaling to CCS [1]:

1. It is often less costly to interface the processing equipment of SPC exchanges with a relatively small number of signaling links than to provide pools of MF registers and line-signaling hardware for the individual trunks.
2. Common-channel signaling is much faster than multifrequency signaling. Early CCS systems already reduced postdialing delays on long-distance calls from 10–15 s to around 3 s.
3. New telecommunication technology and services require the transfer of additional signaling information for processing a call. In Chapter 4 we encountered some additional information items (echo suppressor information, calling line category, etc.). In Signaling System R1 and Signaling System R2 that information is in the form of digits with special meanings (see Section 4.4 and Tables 4.4-4 and 4.4-5). Common-channel signaling messages provide a more flexible way to transfer both the classic supervision and address signals and other types of call-control information.
4. Subscribers cannot access the CCS signaling links. This avoids the "blue-box" fraud problems that have plagued many frequency-division

Signaling in Telecommunication Networks, Second Edition, by John G. van Bosse and Fabrizio U. Devetak

multiplexing (FDM) trunk groups that use channel-associated signaling with in-band signal frequencies (Section 4.4.6).

5. In channel-associated signaling the signals on a trunk necessarily relate to that trunk and are used for call control. In CCS the messages can be—but do not have to be—related to individual trunks. Call control for trunks was the original application of CCS and still is the predominant one. However, CCS links have since become a common transport facility for call control and other applications (Sections 2.1.3–2.1.5).

First-generation common-channel signaling was introduced in the 1970s. It came in two versions: Common-Channel Interoffice Signaling (CCIS), defined by the Bell System, which was deployed in the U.S. network and has since been replaced, and Signaling System No. 6 (also known as CCITT No. 6), an international version defined by CCITT (now ITU-T), which also has been replaced.

CCIS and Signaling System No. 6 (SS6) were followed, about ten years later, by Signaling System No. 7 (SS7). SS7 also exists in several versions. The version specified by ITU-T is widely in service in the international network and, with country-specific modifications, in most national networks. A version defined by the American National Standards Institute (ANSI) and Bellcore (now Telcordia) is in operation in the United States, where it replaced CCIS.

This chapter introduces a number of basic CCS concepts, setting the stage for more detailed discussions of individual CCS systems in later chapters. CCIS and SS6 are not described in this book and for them we refer the reader to [2,3].

5.1 SIGNALING NETWORKS

Telecommunication networks that employ CCS signaling require, in addition to the network of trunks and exchanges, a *signaling network*. This network consists of *signaling points* (SPs), interconnected by *signaling links* (SLs). We start with a few definitions.

Signaling Point (SP). A signaling point is an entity in the network to which CCS signaling links are attached. For example, an exchange that serves CCS trunk groups has CCS signaling links and is therefore a signaling point. Likewise, a network database that is accessed via CCS signaling links is a signaling point.

Signaling Link (SL). A signaling link is a bidirectional transport facility for CCS messages between two signaling points.

Signaling Relation. A signaling relation exists between any pair of signaling points that need to communicate by CCS. For example, when two exchanges, say, A and B, are interconnected by a group of CCS trunks, there is a signaling relation between these signaling points. In what follows we denote a signaling relation between points A and B by (A,B).

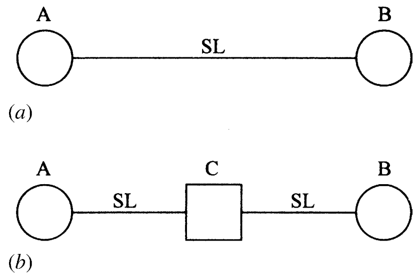


Figure 5.1-1. Signaling routes for relation (A,B): (*a*) associated signaling and (*b*) quasiassociated signaling.

Signaling Route. A signaling route is a predetermined path for the CCS messages of a particular relation. Usually, there is a *signaling route set*, consisting of several routes, for each signaling relation.

Associated and Quasiassociated Signaling. When messages for relation (A,B) are carried on a signaling route that consists of a direct signaling link (SL) between A and B, we speak of *associated* signaling—Fig. 5.1-1(*a*). When A and B have a signaling relation but are not directly interconnected by a SL, a signaling route for relation (*A*,*B*) consists of two or more SLs in tandem—Fig. 5.1-l(*b*). This signaling mode is called *quasiassociated signaling.*

Types of Signaling Points. In Fig. 5.1-1(*a*) and 5.1-1(*b*), signaling points A and B, which originate and receive (and process) signaling messages for relation (A,B), are known as the *signaling end points* (SEPs) for that relation. In Fig. 5.1-1(*b*), signaling point C transfers messages for relation (A,B) but does not originate or process messages for that relation. We say that signaling point C is a *signal transfer point* (STP) for signaling relation (A,B).

Some signaling networks include signaling points that function as end points for some relations and as transfer points for other relations. These dual purpose signaling points are called *signal transfer and end points* (STEPs).

In documents on signaling networks, the various types of signaling points are usually shown as in Fig. 5.1-2. The letter "E" is used in this section to indicate signaling points (SEP or STEP) that are exchanges with CCS trunks.

5.1.1 Basic Signaling Networks

As a start, we explore some alternative signaling networks to handle CCS signaling between exchanges A, B, and C that are interconnected by CCS trunk groups TG_1, TG_2, and TG_3—see Fig. 5.1-3. The signaling network therefore requires routes for relations (A,B), (B,C), and (C,A).

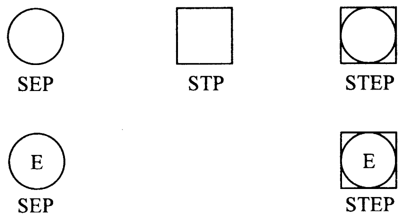


Figure 5.1-2. Signaling points. SEP, Signaling end point; STP, signal transfer point; STEP, combined end/transfer point; E, signaling point and exchange.

Associated Network. Figure 5.1-3(*a*) shows a network in which each route consists of one signaling link that is *associated* with one signaling relation. For example, route SL_1 is associated with signaling relation (A,B) and transports the signaling messages for the trunks in trunk group TG_1 only. All exchanges in the figure are signaling end points.

Signaling links in associated operation are often poorly used, because a link can handle the signaling messages for several thousand trunks, while most trunk groups consist of fewer than 100 trunks.

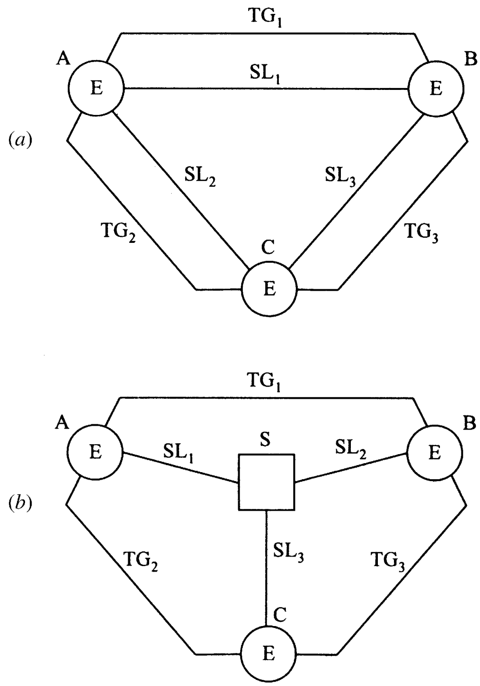


Figure 5.1-3. Signaling networks (*a*) for associated operation and (*b*) for quasiassociated operation.

Quasiassociated Network. Figure 5.1-3(*b*) shows a configuration for quasiassociated operation of the signaling links. None of the exchanges are directly connected by a signaling link. Instead, each exchange has a link to signal transfer point S. Each link carries messages for several relations. All signaling routes are indirect, traversing two links in tandem, and passing through signal transfer point S. All exchanges are again signaling end points.

The structures of associated and quasiassociated signaling networks resemble, respectively, the mesh and star configurations of trunk groups in telecommunications networks.

5.1.2 Signaling Reliability and Load Sharing

In the networks of Fig. 5.1-3 there is one signaling route for each relation. A failure of a signaling link disables the signaling route(s) for which it carries CCS messages, and this severely affects the service in a telecommunications network. For example, a failure of SL_1 in Fig. 5.1-3(*a*) stops all signaling for relation (A,B) and thus shuts down all trunks in trunk group TG_1. Also, on failure of SL_1 in Fig. 5.1-3(*b*), the trunks in groups TG_1 and TG_2 are disabled, and signaling point A becomes isolated.

Actual signaling networks are therefore designed with redundancy, such that signaling for all relations remains possible when a link failure occurs.

Redundancy can be obtained in several ways. For example, the signaling routes in the associated signaling network of Fig. 5.1-3(*a*) can be replaced by route sets containing two direct routes each (see Fig. 5.1-4). Then if, say, SL_1 fails, signaling for relation (A,B) is still possible, using SL_4.

In the quasiassociated configuration of Fig. 5.1-3(*b*), redundancy can be obtained by adding a second signal transfer point with links to each signaling end point. This creates a network in which the route sets for each signaling relation consist of two quasiassociated routes (Fig. 5.1-5). Under normal conditions, the signaling traffic for a trunk group is divided across the signaling routes in a route set. For example, signaling for the odd numbered trunks in TG_1 is on route SL_1-SL_2, and signaling for the even numbered trunks is on route SL_4-SL_5. When SL_1 fails, route SL_1-SL_2 is disabled, and all signaling traffic for TG_1 is carried by the other route.

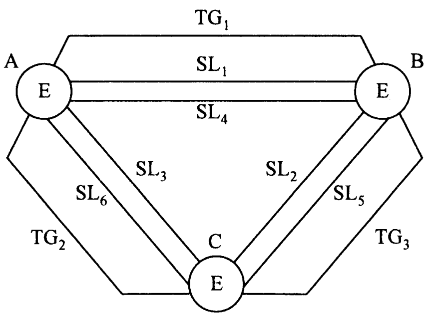


Figure 5.1-4. Signaling network with two routes for each signaling relation (associated signaling).

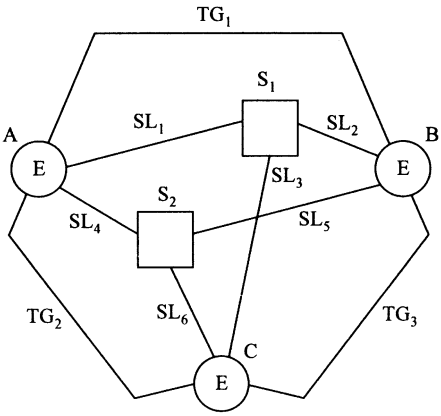


Figure 5.1-5. Signaling network with two routes for each signaling relation (quasiassociated signaling).

A third alternative is shown in Fig. 5.1-6. This arrangement differs from Fig. 5.1-3(*a*) in that exchanges A, B, and C are now *signaling transfer and end points* (STEPs). Under normal conditions, all signaling is associated. However, when, for example, SL_1 fails, the messages related to trunks of group TG_1 are sent via C, which then acts as the STP for the signaling traffic of relation (A,B).

We conclude this section by exploring two well known signaling networks.

5.1.3 The (Former) Bell System Signaling Network

Figure 5.1-7 shows part of the quasiassociated signaling network originally deployed by the Bell System for CCIS [2,4]. The basic structure has been retained for SS7 signaling in AT&T's present long-distance network [5].

The territory of the United States is divided into a number of regions, and each region is equipped with a pair of STPs (only two regions are shown in Fig. 5.1-7). Each exchange with CCS trunks is a signaling end point and has an

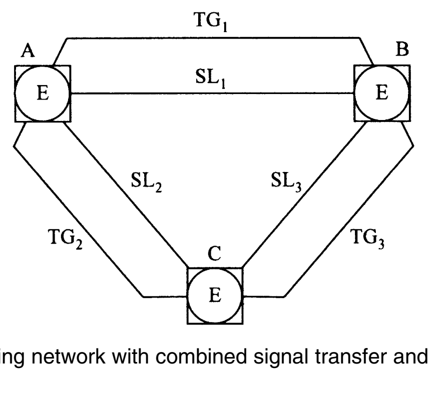


Figure 5.1-6. Signaling network with combined signal transfer and end points (STEPs).

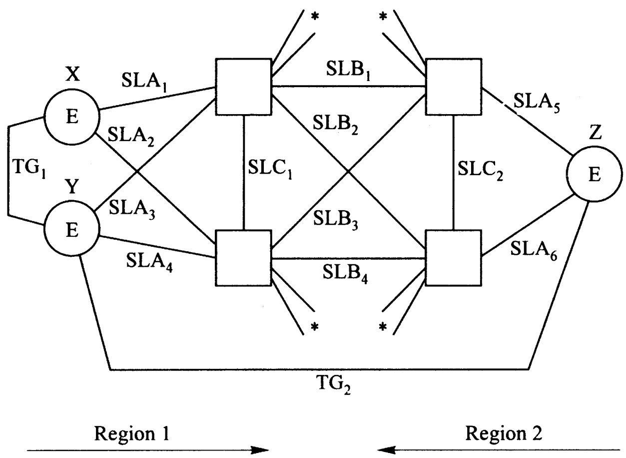


Figure 5.1-7. Bell System signalling network. *Signaling links to other regions. (From *IEEE Commun. Mag.* **28**: 7 Copyright © 1990 IEEE.)

A-link (SLA) to the two STPs in its region. The *B-links* (SLB) interconnect STPs of different regions, and the *C-links* (SLC) interconnect STP pairs of individual regions. The STPs in the network are thus interconnected by a complete mesh of signaling links. The C-links normally do not carry signaling traffic and are used only under certain failure conditions.

In normal operation, the signaling network has two signaling routes R for relations between two SEPs located in the same region. For example, the route set for relation (X,Y) consists of

$$R(X,Y)_1: \quad SLA_1 - SLA_3$$
$$R(X,Y)_2: \quad SLA_2 - SLA_4$$

Also, in normal operation, the signaling route set for a relation between SEPs located in different regions consists of four routes. For example, the route set for relation (Y,Z) consists of

$$R(Y,Z)_1: \quad SLA_3 - SLB_1 - SLA_5$$
$$R(Y,Z)_2: \quad SLA_3 - SLB_2 - SLA_6$$
$$R(Y,Z)_3: \quad SLA_4 - SLB_3 - SLA_5$$
$$R(Y,Z)_4: \quad SLA_4 - SLB_4 - SLA_6$$

In this arrangement, the routes of a route set again share the message load for a signaling relation. For example, trunk group TG_2 can be divided (for signaling purposes only) into four subgroups. When all routes for relation (Y,Z) are operational, the message traffic for a particular subgroup is carried by one of the four routes.

When a route is disabled, the signaling traffic for the affected subgroup is diverted to one of the remaining routes.

Simultaneous failures of signaling links in a route set can disable all normal signaling routes for a relation. In Fig. 5.1-7, simultaneous failures of SLA_1 and SLA_4 disable both signaling routes for (X,Y). In this case, signaling is maintained by using the route SLA_2–SLC_1–SLA_3, which does not belong to the normal route set for (X,Y).

5.1.4 Mixed Signaling Networks

Some networks use associated signaling for some relations and quasiassociated signaling for the others relations. Figure 5.1-8 shows an example that is often used between international exchanges (ISCs) in two countries. Exchanges X and Y in country A are connected to exchanges Z and U in country B by international CCS trunk groups $TG_1, \ldots, TG_4$, and by two international signaling links (SL_1,SL_2). Each pair of in-country exchanges is also interconnected by national signaling link (SL_3,SL_4). All exchanges are combined signal transfer and end points.

The normal (nonfailure) signaling routes can be assigned as follows:

R(X,Z):	SL_1	(associated signaling)
R(Y,U):	SL_2	(associated signaling)
R(X,U):	SL_3–SL_2	(nonassociated signaling)
R(Y,Z):	SL_3–SL_1	(nonassociated signaling)

In this example, there is only one normal route for each signaling relation. However, signaling for all relations can be maintained on failure of a signaling link, by using routes that are not used for signaling relations under normal conditions. For example, when SL_1 fails, the message traffic for relation (X,Z) can be diverted to the route SL_3–SL_2–SL_4, and the traffic for relation (Y,Z) can be diverted to the route SL_2–SL_4.

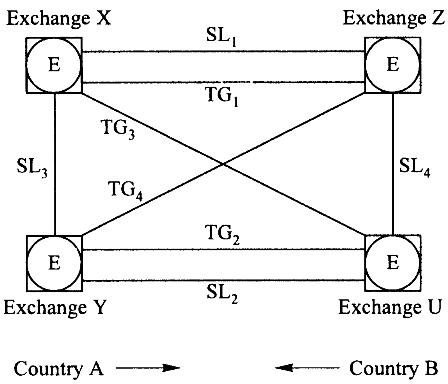


Figure 5.1-8. Mixed-mode signaling network.

5.2 SIGNALING LINKS AND SIGNAL UNITS

This section introduces some fundamental aspects of CCS signaling links and signal units. The signaling links are described using the hardware-oriented terms of the literature on SS6 [2,3,5]. The characteristics of the signaling links in SS7 are quite similar but are described in a more abstract manner (see Chapter 8).

5.2.1 Signaling Link

Figure 5.2-1 shows a signaling link between signaling points A and B. It consists of two *signaling terminals* (STs) and a bidirectional *signaling data link* that transfers digital data. The primary function of the signaling link is to provide a reliable transfer of signaling messages between processors P_A and P_B.

In a signaling point, there are four interfaces between the processor and a signaling terminal. The processor enters its outgoing messages M_O into—and retrieves its incoming messages M_I from—the ST. In addition, the processor can send commands (COM) to the ST—for example, to activate or deactivate the link. A ST sends indications (IND) to alert the processor about certain conditions on the link (excessive errors in received messages, overload, etc.).

5.2.2 Signal Units

Information is transferred across the signaling data link in *signal units* (SUs) (groups of consecutive bits). We distinguish two SU types—see Fig. 5.2-2. *Message* SUs (a) transfer processor messages (b). *Link* SUs (c) transfer information originated by the ST at one end of the signaling link and intended for the ST at the other end.

A processor message is an ordered set of digitally coded *parameters* (information elements). The initial parameters identify the message type and imply the meanings and locations of the later parameters in the message.

The length (number of bits) of SUs is variable. The contents of SUs are shown in Fig. 5.2-3. All SUs have an information field (INF) and a check-bit field (CB). The INF field of a message SU holds all parameters of a message, and several link parameters.

The INF fields of link SUs contain link parameters only.

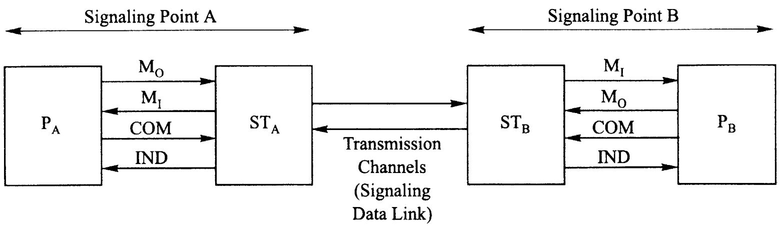


Figure 5.2-1. Signaling link.

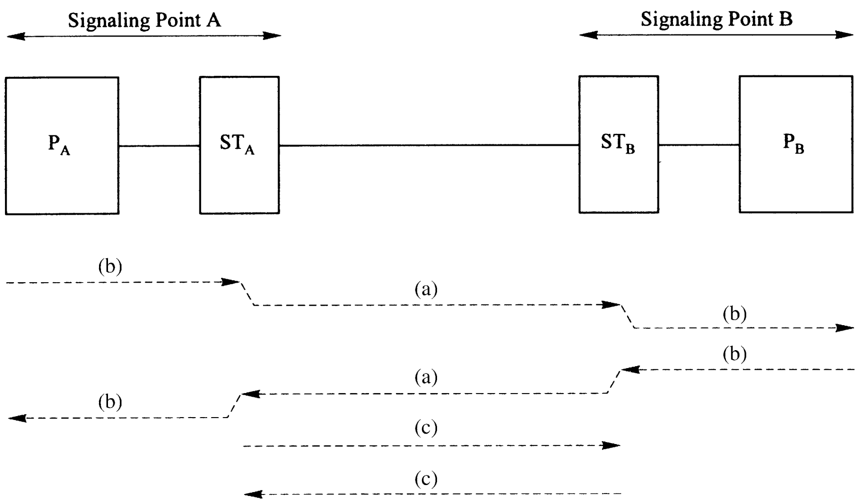


Figure 5.2-2. Signal units and processor messages. (a), Message signal unit; (b), processor message; (c), link signal unit.

5.2.3 Signaling Terminal Functions

The most important signaling terminal functions are synchronization, the transfer of SUs (including error control for message SUs), and monitoring of the signaling link.

Synchronization. A working signaling link conveys an uninterrupted bit stream of adjacent SUs in each direction. When no message SU needs to be transmitted, a ST sends certain types of link SUs (*filler* or *synchronization* SUs). The "receive" part of a ST has to be *synchronized* (*aligned*) with its incoming bit stream, so that it can determine the start points of the individual SUs. Synchronization needs

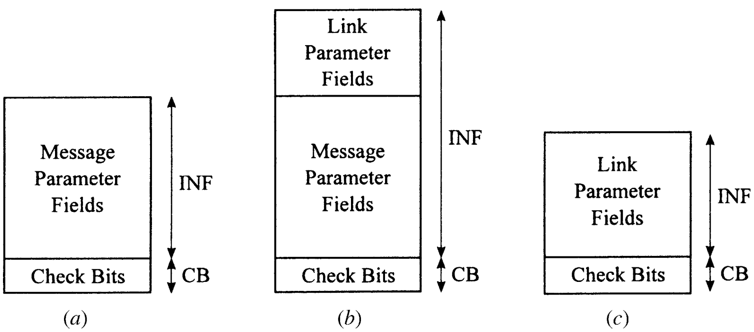


Figure 5.2-3. Contents of signal units: (*a*) SS6 message signal unit, (*b*) SS7 message signal unit, and (*c*) SS6 and SS7 link signal unit.

to be acquired when a link is turned "on" and reacquired when a disturbance on the link causes a terminal to lose its alignment. In SS6, alignment and realignment procedures rely on a specific bit pattern in synchronization SUs. In SS7, adjacent SUs are separated by "flags" that have a bit pattern that does not occur in SUs.

Error Control. Electrical disturbances on a signaling link can introduce errors in signal units, changing one or more 1's into 0's, and vice versa. Errors in message SUs can cause more serious problems than errors in the signals of channel-associated system. For example, errors in a particular call-control message for a particular trunk can change the message into a message for another trunk, or into a different message for the intended trunk, and thus cause the processor to take a wrong call-control action.

In addition, the signaling speed on data links is on the order of several thousand bits per second. A disturbance on a signaling link of just a few milliseconds can therefore severely mutilate a SU, while a similar disturbance on a trunk with channel-associated signaling (in which signal recognition times are on the order of 50–200 ms) is usually harmless.

Signaling terminals therefore execute error control procedures when sending and receiving SUs. The procedure for SUs sent by ST_A to ST_B is described below (the same holds for transmission in the other direction).

Error Detection. Error detection makes use of the CB fields of SUs. The contents of the CB fields are calculated by ST_A just before the SU is sent and enable ST_B to determine whether a received SU is error-free.

Acceptance and Acknowledgment of Message SUs. ST_B determines whether a received SU is free of errors. Error-free SUs are accepted, and SUs with errors are discarded. Moreover, ST_B "positively" acknowledges error-free message SUs and "negatively" acknowledges message SUs received with errors. In SS6, ST_B sends acknowledgments in special link SUs (acknowledgment SUs). In SS7, ST_B sends acknowledgment information by including link parameters in message and link SUs it sends to ST_A. Received link SUs are not acknowledged.

Retransmission of Message SUs. When ST_A receives a negative acknowledgment of a sent message SU, it retransmits the SU.

Monitoring. The signaling terminals at both ends of a signaling link monitor the condition of the link by keeping track of alignment losses, SUs received with errors, and so on. When a misalignment lasts too long, or when the fraction of SUs received with errors becomes too high, the ST alerts its processor with an indication (IND) (Fig. 5.2-1). The processor may then decide to take the signaling link out of service.

5.2.4 Signaling Terminal Elements and Operation

The major functional elements of a signaling terminal are shown in Fig. 5.2-4. *Terminal control* (TC) controls the ST, originates the link parameters of outgoing SUs, and processes the link parameters of incoming SUs.

The processor at a signaling point places its outgoing messages M_o in *output buffer* (OB) and retrieves incoming messages M_I from *input buffer* (IB). *Retransmission buffer* (RB) stores messages that have been sent out but have not yet been acknowledged positively.

Sending SUs. Information to be sent out can come from three sources: output buffer (OB), retransmission buffer (RB), or terminal control (TC) (link information).

After a SU has been sent, the selection of the next SU to be sent is made according to the following priority scheme:

Priority 1: A link SU, if TC needs to alert the distant TC about an event on the link.

Priority 2: Retransmission of a message in RB that is marked as "to be retransmitted."

Priority 3: Initial transmission of a message in OB.

Priority 4: A "filler" link SU (if nothing else is waiting to be sent).

Initial Transmission of a Message in OB. The message is removed from OB and is entered into RB and into *output processing* (OP). The SU is assembled in OP, which first forms the INF part of the SU (in SS7, this involves adding the link parameters to the message parameters). OP then calculates the contents of CB and appends it to INF.

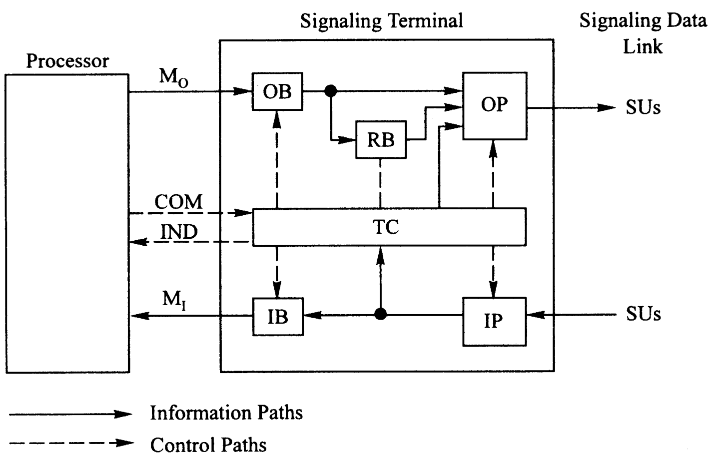


Figure 5.2-4. Signaling terminal.

Retransmission of a Message in RB. A copy of the message is entered into OP, where the SU is formed as described above. The message itself remains in RB.

Receiving a Signal Unit. All received SUs are checked for errors by *input processing* (IP). SUs with errors are discarded. Error-free SUs are processed further. If the SU is a message SU, its message parameters, which constitute the processor message, are entered into buffer IB. In SS7, the link parameters of a message SU are passed to TC. In SS6 and SS7, the parameters in link SUs are passed to TC. Some link parameters acknowledge received message SUs. When TC receives a positive acknowledgment of a message in RB, the message is deleted. When TC receives a negative acknowledgment of a message in RB, the message is marked as "to be retransmitted."

5.2.5 Message Sequencing

We have seen that the retransmission of messages in buffer RB has priority over the initial transmission of messages in OB. We now examine the selection for transmission of a message in a buffer.

CCS messages belong to several signaling applications, such as call-control or management of trunks. Each application has a priority class. Buffers OB and RB are operated "first-in, first out" for each application: the message selected for transmission is the "oldest" message with the highest priority class.

Let us focus two call-control messages, M_1 and M_2, in buffer OB of signaling terminal ST_A (Fig. 5.2-5). M_1 was entered in buffer OB by processor P_A before message M_2.

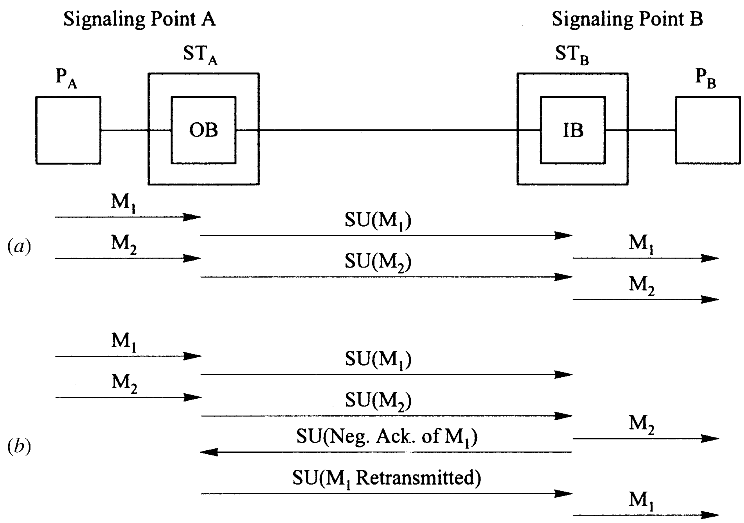


Figure 5.2-5. Transfer of call-control messages M_1 and M_2.

Since M_1 is "older," its initial transmission takes place before the initial transmission of M_2. Assuming that both messages are received error-free and accepted by ST_B, M_1 is entered in buffer IB of ST_B before M_2. Buffer IB is also operated "first-in, first out" for each application priority, and processor P_B retrieves and processes M_1 before M_2—Fig. 5.2-5(*a*).

We speak here of "in-sequence" delivery of messages. In-sequence delivery is of importance when messages M_1 and M_2 pertain to the same call.

Now suppose that message M_1 is received with errors by ST_B and therefore is discarded—Fig. 5.2-5(*b*). ST_B informs ST_A by sending a *negative* acknowledgment of M_1.

In SS6, ST_A retransmits message M_1 only, and processor P_B therefore receives M_1 *after* M_2—Fig. 5.2-5(*a*). When these messages pertain to the same call, this out-of-sequence (O-S) delivery causes a problem for P_B in the processing of the call.

In SS7, when signaling terminal ST_B receives message M_1 with errors, it discards the message—and all following messages, until it receives an error-free retransmission of M_1. ST_A, after receiving a negative acknowledgment of message M_1, retransmits all sent messages (starting with M_1) before sending out any new message. In this way, SS7 eliminates the major cause of O-S message deliveries.

We shall see later that certain failures of signaling links can also result in O-S deliveries. SS7 thus reduces the probability of such deliveries but does not eliminate them. Therefore, the call-control procedures in both SS6 and SS7 have to take into account the possibility that messages are not delivered in their proper sequence.

5.2.6 Cyclic Redundancy Checking

Error detection of SUs in SS6 and SS7 is done by *cyclic redundancy checking* (CRC), a technique used in many data communication systems. The CRC procedure for SUs from ST_A to ST_B is described below (Fig. 5.2-6).

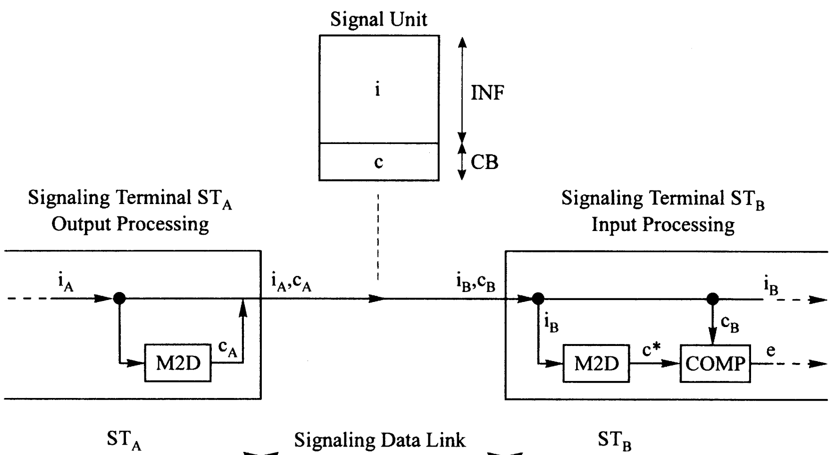


Figure 5.2-6. Cyclic redundancy checking.

In the final step of output processing at ST_A, the contents of INF is regarded as a binary number (i_A). This number is fed into a *modulo 2 divider* (M2D), where it is divided (mod 2) by a divisor d. The remainder of the division (c_A) is entered into CB of the SU.

Suppose now that the SU arrives at ST_B, and that the contents of INF and CB are i_B and c_B. In the initial step of input processing at ST_B, i_B is fed into M2D, where it is divided (mod 2) by the divisor d mentioned above. The division yields a remainder c^*.

Comparison circuit COMP compares c_B and c^*. If there are no errors in INF ($i_B = i_A$), then $c^* = c_A$. If there also are no errors in CB ($c_B = c_A$), the inputs to COMP are equal, and its output $e = 0$. This is taken as the indication that the SU is error-free. Otherwise, the SU is deemed to contain errors and is discarded.

Cyclic redundancy checking does not catch all transmission errors but, with a properly selected divisor d, the probability of an undetected SU error can be made very low (on the order of 1 in 10^8).

Modulo 2 Division. Mod 2 arithmetic is binary arithmetic without "carries" or "borrows." For example, consider the mod 2 subtraction (or addition)

$$\begin{array}{l} 101101 \\ \underline{111001} - (\text{or} +) \\ 010100 \end{array}$$

The result is obtained by performing "exclusive or" operations on the bits in corresponding positions of both numbers.

Mod 2 division consists of mod 2 subtractions and shifts. The example below shows the calculation of the remainder c that results from mod 2 division of $i = 1101011$ by divisor $d = 1011$. By long division we have

$$\begin{array}{lrrl}
 & 1011 & \overline{)1101011} & \\
 & & \underline{1011} & \leftarrow \text{mod 2 subtraction} \\
 & & 1100 & \\
 & & \underline{1011} & \leftarrow \text{mod 2 subtraction} \\
 & & 1111 & \\
 & & \underline{1011} & \leftarrow \text{mod 2 subtraction} \\
 & & 1001 & \\
 & & \underline{1011} & \leftarrow \text{mod 2 subtraction} \\
\text{Remainder c:} & & 010 &
\end{array}$$

For more information on CRC, including probabilities of undetected errors, the reader is referred to [6,7].

5.3 ACRONYMS

ANI	Automatic number identification
ANSI	American National Standards Institute

CAS	Channel-associated signaling
CB	Check-bit field of signal unit
CCIS	Common-Channel Interoffice Signaling
CCITT	International Telegraph and Telephone Consultative Committee
CCS	Common-channel signaling
COM	Command
COMP	Comparison circuit
CRC	Cyclic redundancy checking
FDM	Frequency-division multiplexing
IB	Input buffer
IND	Indication
INF	Information field of signal unit
IP	Input processing
ISC	International switching center
ITU	International Telecommunication Union
ITU-T	Telecommunication Standardization Sector of ITU
M	Message
M2D	Modulo 2 divider
MF	Multifrequency
OB	Output buffer
OP	Output processing
P	Processor
RB	Retransmission buffer
SEP	Signaling end point
SL	Signaling link
SP	Signaling point
SPC	Stored-program controlled
SS6	Signaling System No. 6
SS7	Signaling System No. 7
ST	Signaling terminal
STEP	Signaling transfer and end point
STP	Signal transfer point
SU	Signal unit
TC	Terminal control
TG	Trunk group

5.4 REFERENCES

1. *A History of Engineering and Science in the Bell System*, Chapter 12, Bell Telephone Laboratories, 1982.

2. C. A. Dahlbom and J. S. Ryan, History and Description of a New Signaling System, *Bell Sys. Tech. J.*, **57**, 1978.
3. Specification of Signaling System No. 6, Rec. Q.251–Q.300, ITU-T, Geneva, 1988.
4. *Notes on the Network*, AT&T, New York, 1980.
5. A. R. Modaressi and R. A. Skoog, Signaling System No. 7: A Tutorial, *IEEE Commun. Mag*., July 1990.
6. W. Stallings, *ISDN*, Macmillan, New York, 1989.
7. W. Peterson and D. Brown, Cyclic Codes for Error Detection, *Proc. IRE*, Jan. 1961.

6

SIGNALING IN ACCESS NETWORKS

Section 1.8 introduced access networks and showed how access systems (ASs) present a standard physical interface (e.g., E1 or T1) as well as a standard signaling interface to the local exchange (LE). Standard interfaces allow access systems and local exchanges to be procured from different manufacturers without creating inter-working problems. This section discusses signaling protocols between access systems and local exchanges.

6.1 OVERVIEW OF SIGNALING FOR ACCESS SYSTEMS

Interface standards define an access system as a unit that serves a configurable number of subscriber lines and connects to a local exchange via a configurable number of digital links. The maximum number of digital links in a given unit is a parameter specified by the interface standards and determines the traffic capacity of the unit. If additional capacity is required, additional units must be deployed. The digital links in one unit are referred to collectively as an "AS interface."

With access systems, responsibility for call control always resides in the local exchange, and no intra-access-system switching is supported.

The main AS functions that involve signaling are:

- Time-slot (TS) management
- Call control for analog lines
- Protection switching

Signaling in Telecommunication Networks, Second Edition, by John G. van Bosse and Fabrizio U. Devetak

Time-Slot Management. This function assigns and/or activates time slots in a digital link for connections to analog or ISDN lines, then releases and/or deactivates them when they are no longer needed. With statistical multiplexing (Section 1.8), signaling handles the dynamic assignment of time slots to lines at the start of a call and their release at the end of it. With fixed (semipermanent) assignment, signaling handles the activation and deactivation of preset connections of time slots to lines.

Call Control for Analog Lines. This function starts after a line has been assigned a time slot and the connection between line and TS has been activated. Subscriber loop signals from the subscriber are converted to digital signals that can be carried by the digital link; digital signals from the LE are converted to analog signals that can be applied toward the subscriber. DTMF signals from subscribers and audible signals from the LE (such as dial-tone and busy-tone) go through the AS transparently, like voice.

Call-control signaling for analog lines is based on *stimulus protocols*. Most signaling protocols, both common-channel and channel-associated, are functional protocols, requiring equal processing complexity in sender and receiver. Stimulus protocols, on the other hand, use simple commands that require only one end to keep track of call states, thus keeping complexity at the other end to a minimum. In access networks, stimulus signaling keeps call processing simple in the AS and concentrates complexity in the LE. Signaling from the AS to the LE consists of indications of changes detected in the electrical configuration of the subscriber line (e.g., open or closed loop). Signaling from the LE to the AS consists of instructions to deploy electrical configurations (e.g., ringing current or reverse polarity) toward the subscriber line.

Protection Switching. This function is the ability to maintain communication between the AS and the LE when a digital link fails. Protection is realized by configuring active (protected) and standby (protecting) channels. Obviously, this function applies only to multilink configurations, and protecting and protected channels must be in different links.

Call Control for ISDN Lines. In addition to analog lines, ASs support digital subscriber lines that use ISDN (DSS1) signaling (Chapter 10). ISDN call-control signaling is carried in D-channel messages, which are transported in time slots assigned semipermanently by provisioning and transit through the AS with little or no involvement by AS signaling.

Signaling Protocols. The two most commonly used AS protocols are:

- GR-303, issued by Telcordia
- V5.1 and V5.2, issued by ITU-T and ETSI

TABLE 6.1-1 Access Systems Terminology Map

Common Term	GR-303	V5
Access system (AS)	Remote digital terminal (RDT)	Access network (AN)
Local exchange (LE)	Local digital switch (LDS)	Local exchange
Digital trunk interface in the LE	Integrated digital terminal (IDT)	Exchange termination (ET)
The complex of AS, transmission facilities, and LE terminations	Integrated digital loop carrier (IDLC)	(not used)
An individual TS in a DS1 line	DS0	64-kb/s Time slot
Time-slot assignment	Time-slot assignment	Time-slot allocation
Voice channel[a]	TS assigned to a subscriber line	Bearer channel

[a]A voice channel can carry user data other than voice, but voice is the most common use.

Both protocols use the concepts of *common-channel signaling* (Chapter 5) and *layers* (Chapter 20), which are necessary for a full understanding of the sections that follow. The reader unfamiliar with those concepts is referred to the corresponding chapters of the book.

Terminology. The terminology used in the Telcordia interface standards is different from the terminology used in the ITU-T recommendations. For ease of understanding, this chapter uses the standard telecommunications terms found in the rest of the book. Table 6.1-1 provides a map of the different terms.

6.2 THE GR-303 STANDARD

GR-303 (formerly TR-303) is the North American standard for access interfaces [1]. The standard supports:

- Both fixed and dynamic TS assignment
- Analog subscriber lines
- ISDN subscriber lines (2B + D and 23B + D)
- 1 to 28 T1/DS1 digital links per GR-303 interface
- SONET optical links
- Protection switching for multilink configurations

Channels. Most time slots (DS0) in a digital link are used as voice channels that can be assigned to subscriber interfaces; a few are used as signaling and OAM channels. There are two types of signaling channels:

- Common signaling channel
- Channel-associated signaling subchannels

Common Signaling Channel (TS 24). This is a DS0 channel used for TS assignment and optionally to carry call-control signaling for analog lines. When used only for TS assignment, it is called the Timeslot Management Channel (TMC); when used also for call control it is called the Common Signaling Channel (CSC). There is one active common signaling channel per GR-303 interface. A standby channel is also present in multiple-DS1 configurations (see later under "path protection").

Channel-Associated Signaling Subchannels. They carry call-control signaling for analog lines when the TMC is used for TS assignment. Each TS used as a voice channel has its own signaling subchannel, obtained by "robbing" the least significant bit in frames 6, 12, 18, and 24. That is possible because GR-303 interfaces use the *extended superframe* (ESF) format (Section 4.2.3). The four bits are called A, B, C, D. This configuration is called *signaling option 16* because of the 16 signaling states (2^4). If a SONET (optical fiber) interface is used, the ABCD bits are carried in the overhead part of SONET frames; so eight bits are available for user traffic in every frame, providing a 64-kb/s *clear channel connection* (CCC). The signaling configuration in this case is called *signaling option T* (transparent).

With the CSC, call-control signaling is transported out-of-band so, regardless of the transmission medium, a voice channel always provides a clear channel connection.

OAM (Operations) Channels. They are the following:

- Embedded Operations Channel (EOC)
- Loop testing channels
- ESF data link

The EOC is a DS0 channel (TS 12), one per GR-303 interface, used for operations messages exchanged between the AS and the LE and between the AS and an external OS. In the latter case the LE relays messages between the two.

Loop testing channels are DS0 channels, one or more per interface.

The ESF data link is a 4-kb/s data link, one per GR-303 interface, derived from the 193rd bit of every DS1 frame (Section 4.2.3). It is used for DS1 facility protection messages (see below) and for facility performance reports.

Protocols used on the OAM channels are not discussed in this book and readers interested in them are referred to [1].

Protection Switching. This comes in two forms:

1. *Path protection switching* for the TMC/CSC and the EOC (mandatory for multilink configurations).
2. *Facility protection switching* for DS1 facilities (optional) and for optical facilities (mandatory). Optical facilities must always have at least two links.

For path protection, the TMC/CSC and the EOC each have an active and a standby channel. When path protection is triggered, traffic is switched from the active to the standby channel. Switching is managed via operations messages on both active and standby channels.

Facility protection switching uses a $n + 1$ scheme and is managed via messages on the ESF link.

Protection switching can be manual or automatic. Manual switching can be initiated by an operations system (OS) or by the human–machine interface at the LE or AS. Automatic switching is initiated by the LE or AS, depending on which side detects failure or deterioration in the protected resource.

Signaling. GR-303 supports two signaling methods:

- Hybrid signaling
- Common Signaling Channel (CSC) signaling

Support of hybrid signaling is mandatory; support of CSC signaling is optional. A signaling method applies to the whole GR-303 interface and cannot be configured on a per-line basis.

For analog lines each signaling method manages both TS assignment and call control. Hybrid signaling uses the TMC for TS assignment and the channel-associated signaling subchannels for call control. CSC signaling (also called *out-of-band signaling*) uses the CSC for both TS assignment and call control.

ISDN lines use only the TS assignment function of each signaling method: the TSC protocol with hybrid signaling and the TS assignment part of the CSC protocol with CSC signaling. The call-control part of the signaling method is not used, since call control is carried in D-channel messages (Section 6.2.4).

Signaling on the TMC and the CSC is out-of-band and message-oriented; that is, it is a form of common-channel signaling (Chapter 5) with messages carried in a dedicated TS. The protocol stack is shown in Fig. 6.2-1.

6.2.1 Data Link Layer Protocol

The data link layer protocol for the TMC and the CSC is a subset of DSS1 LAPD (Section 10.2 and [2]), which uses the multiframe transfer mode. The subset of LAPD Command/Response Frames is listed in Table 6.2-1.

Application Layer	TMC Signaling	CSC Signaling
Data Link Layer	LAPD (Q.921)	
Physical Layer	T1 or SONET	

Figure 6.2-1. Protocol stack for TMC and CSC signaling.

TABLE 6.2-1 LAPD Frame Types Used by GR-303

Frame Type	Full Name
I	Information
SABME	Set Automatic Balance Mode Extended
RR	Receive Ready
REJ	Reject
RNR	Receive Not Ready
UA	Unnumbered Acknowledgment
DM	Disconnected Mode

Call processing and TS assignment messages use SAPI = 0. Path protection messages use SAPI = 1. SAPI is discussed in Section 10.2.

6.2.2 Hybrid Signaling for Analog Lines

Time-slot assignment, handled by messages on the TMC, uses the Q.931 ISDN protocol (Section 10.3), with the general message format of Fig. 10.3-2 and the value of the Protocol Discriminator information element (IE) set to 01001111. Given the limited functionality, a subset of the Q.931 messages is used. The most significant ones are listed in Table 6.2-2.

Time slots are always assigned to a call by the LE. TS identification is contained in the Channel Identification IE, which is carried in the Setup message for calls originated by the LE, and in the Connect message for calls originated by the subscriber. Individual subscriber lines are identified by the Call Reference IE, which is a sequential line number and is included in every TMC message. The LE converts directory numbers into call reference numbers and vice versa. The TS release sequence is always started by a Disconnect message from the LE, even when the AS side hangs up first. In the latter case, on-hook is detected by the LE via ABCD signaling. The sequence is completed with Release followed by Release Complete.

TABLE 6.2-2 Q.931 Messages for TMC and CSC Signaling

TMC Messages	AS ↔ LE	CSC Messages	AS ↔ LE
Connect	→ ←	Alerting	→
Connect Acknowledgment	→	Call Proceeding	→
Disconnect	→ ←	Connect	→ ←
Information	→ ←	Disconnect	→ ←
Release	→ ←	Information	→ ←
Release Complete	→ ←	Release	→ ←
Setup	→ ←	Release Complete	→ ←
		Setup	→ ←
		Setup Acknowledgment	→ ←

Note: CSC messages not shown in the TMC table are not used on TMC.

TABLE 6.2-3 Hybrid Signaling: ABCD Codes for Loop Start Lines

ABCD Code	Signal	Function	AS ↔ LE
0000	-R Ringing	Alert called party	←
0010	DS0 AIS	Alarm indicator signal	→ ←
0100	RCLF	Reverse loop current feed	←
0101	LCF	Apply loop current feed	←
0101	LO	Loop open	→
0111	DS0 Yellow	RAI[a] alarm	→ ←
1111	LCFO	Open loop current feed	←
1111	LC	Loop closed	→

[a]RAI, Remote Alarm Indication.

Call control uses the channel-associated signaling subchannels and supports five types of analog lines: Loop Start, Ground Start, Loop Reverse Battery, Coin, and Multiparty. Table 6.2-3 shows the coding of the ABCD bits for Loop Start lines. Dial pulses are transmitted as a sequence of open loop (LO) and closed loop (LC) indications.

An example of a call sequence for a LE origination is shown in Fig. 6.2-2. On receipt of Setup (TMC), the AS acknowledges with a Connect message (TMC)

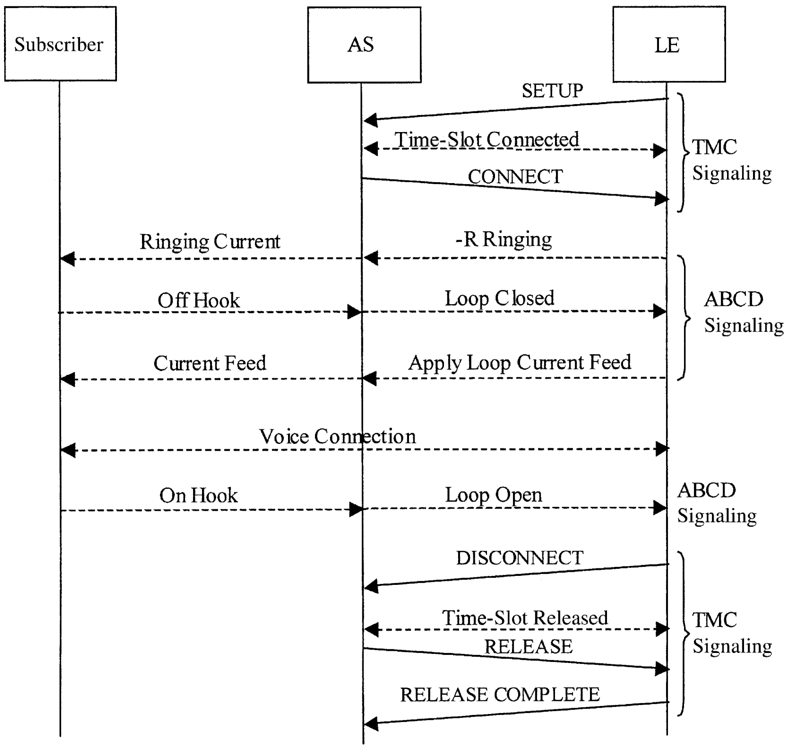


Figure 6.2-2. Hybrid Signaling: LE-originated connection and release for Loop Start lines.

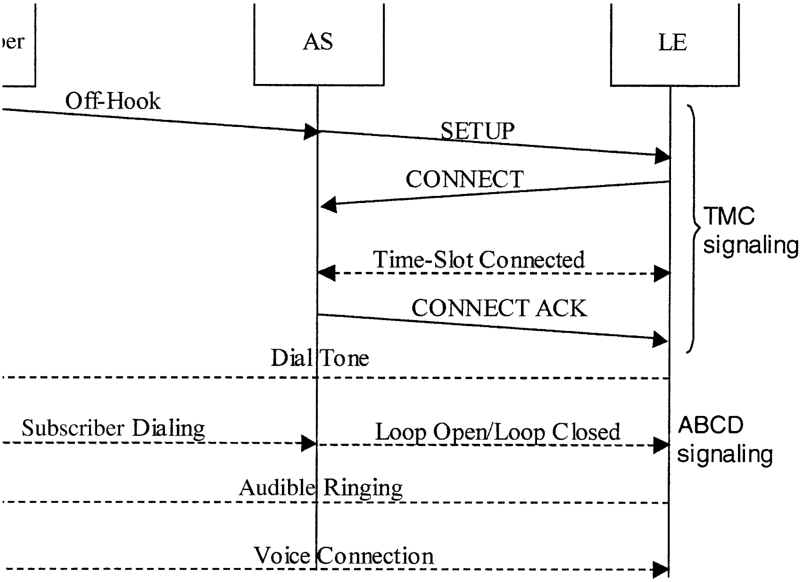


Figure 6.2-3. Hybrid Signaling: AS-originated connection for Loop Start lines.

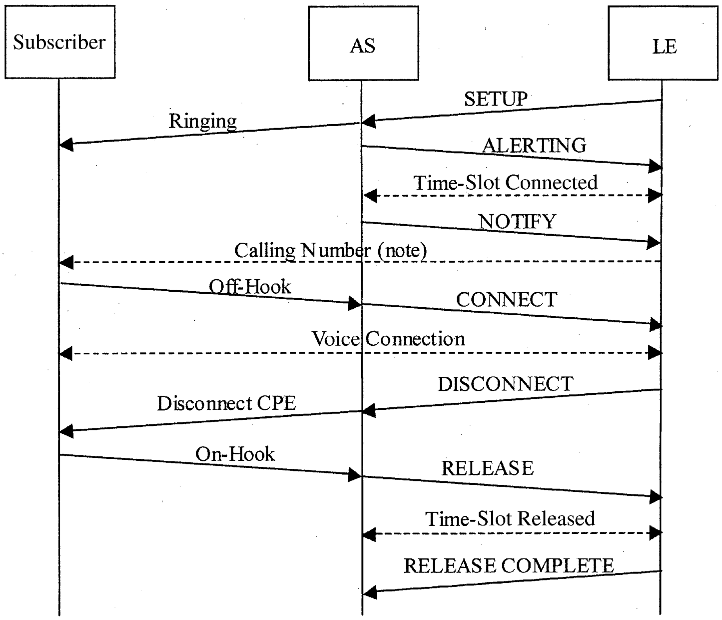
and starts to monitor the allocated time slot for ABCD signals. The LE then sends Ringing (ABCD). When the subscriber answers, AS sends Loop Closed (ABCD). The LE responds with Apply Current Feed (ABCD) and the talking connection is established. If the subscriber hangs up first, the LE receives Loop Closed (ABCD) and sends a Disconnect (TMC) to release the time slot. If the LE side hangs up first, the LE sends a Disconnect (TMC). The release sequence is completed by a Release and Release Complete exchange (both on TMC). Figure 6.2-3 shows an example of an AS origination (the release sequence is the same as in Fig. 6.2-2).

6.2.3 CSC Signaling for Analog Lines

The CSC uses the Q.931 ISDN protocol (Section 10.3) for TS assignment and for call control. The general message format is shown in Fig. 10.3-2 and the value for the Protocol Discriminator IE is 01001111, the same as for TMC messages. A subset of the Q.931 messages is used, and the most significant ones are listed in Table 6.2-2.

TS assignment is controlled by the LE and is similar to the procedure for hybrid signaling. TS identification is contained in the Channel Identification IE, which is carried in the Setup message for LE originations, and in the Setup Acknowledgment message for subscriber originations. The Call Reference IE, included in every CSC message, is used for line identification.

The same five types of analog lines are supported as in hybrid signaling. Dial-pulse information is carried in Info messages (one per digit). Connect is used



Note: optional "on-hook signaling" to transmit Calling Number ID during the pause between ringing pulses.
CPE: Customer Premises Equipment

Figure 6.2-4. CSC Signaling: LE-originated connection and release for Loop Start lines.

by the LE to notify the AS that enough digits have been received and further digits will be ignored.

The release sequence is started by a Disconnect message, which can be sent by the LE or the AS, depending on which side hangs up first. When the AS disconnects first, the LE responds with Notify followed by Release, which the AS acknowledges with Release Complete. When the LE disconnects first, the AS responds with Release, acknowledged by the LE with Release Complete.

Two examples of call sequences with CSC signaling for Loop Start lines are shown in Figs. 6.2-4 and 6.2-5.

6.2.4 Signaling for ISDN Lines

Call-control signaling travels through the AS in D-channel messages with no involvement by GR-303 signaling, as described below. B-channels are always clear.

Basic Access Lines (2B + D). These lines can be configured in two ways.

3-DS0 Method. Each B-channel and each D-channel is connected semipermanently to a separate TS. All TSs are assigned by provisioning and no signaling for TS assignment is necessary. This method is used only for foreign exchange lines

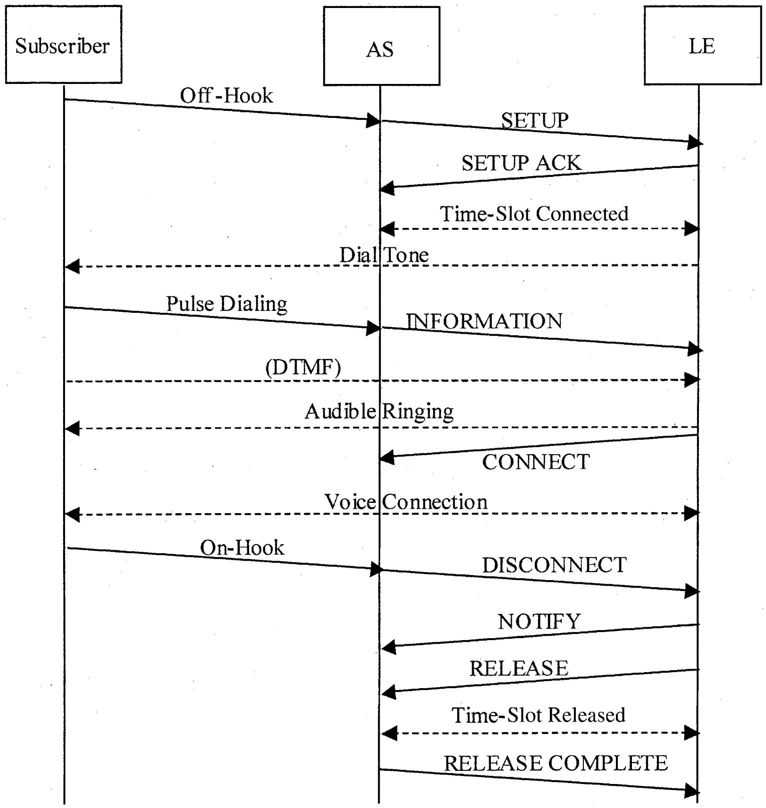


Figure 6.2-5. CSC Signaling: AS-originated connection and release for Loop Start lines.

or lines connected to a cross-connect system; it is supported by the AS only (not by the LE).

4:1 TDM Method. Each B-channel is connected to a separate TS and the D-channels of up to four ISDN lines share one TS. D-channels are assigned to TSs semipermanently and B-channels can be assigned semipermanently or dynamically. This is the standard method for ISDN lines that terminate at the LE hosting the AS.

Dynamic assignment is always started by the LE and line originations are detected by the LE via D-channel signaling. With the hybrid method, the assignment procedure is the same as for analog lines. With the CSC method, the alerting sequence for LE originations is not needed because alerting is handled by D-channel signaling: Setup is sent from the LE with "Alerting Off" in the Signal IE and is followed by Connect from the AS. Clearing of B-channels is always started from the LE and works as shown in Fig. 6.2-4. The call control portion of the CSC protocol is not used.

Primary Rate ISDN Interfaces (23B + D). These interfaces are DS1 lines and each of the 23 B-channels and the D-channel are connected semipermanently to a separate TS by provisioning.

6.3 THE V5 STANDARDS

The V5 standards are named after the *V5 reference point* defined by ITU-T [3]. Reference points are conceptual interfaces between two functional entities and V5 identifies the interface between a digital cross-connect system or a concentrator (such as an AS) and a local exchange.

Two protocol standards, V5.1 and V5.2, have been defined by ITU-T, and ETSI has issued virtually identical standards. The standards documents are listed in Table 6.3-1. V5.1 was the first protocol to be published; V5.2, introduced later, added significant enhancements. The main features of V5.1 and V5.2 are shown in Table 6.3-2. V5 standards apply to 2.048-Mb/s signals (E1 or SDH/STM), in electrical or optical interfaces.

V5 signaling is strictly common-channel, that is, out-of-band, message-oriented, and carried in dedicated time slots. The V5 protocols are structured along three layers: layer 1 (L1) or physical layer, layer 2 (L2) or data link layer, and layer 3 (L3) or network layer.

Time slots in V5 interfaces are used as voice (*bearer*) channels, which carry user traffic, and as communication channels, called *C-channels*, which carry signaling and operations messages. In a C-channel, signaling and operations messages of the same class (regardless of the user interface they relate to) are grouped into

TABLE 6.3-1 V5 Standard Documents

Protocol	ETSI	ITU-T
V5.1	ETSI 300-324-1	G.964
V5.2	ETSI 300-347-1	G.965

TABLE 6.3-2 V5.1 versus V5.2

Supported Function	V5.1	V5.2
Analog lines	Yes	Yes
ISDN basic access (2B + D)	Yes	Yes
ISDN primary rate access (30B + D)	No	Yes
ISDN H-Channels	No	Yes
Leased lines	Yes	Yes
Fixed bearer (voice) channel assignment to lines	Yes	Yes
Dynamic bearer (voice) channel assignment to lines (statistical multiplexing)	No	Yes
Protection switching (backup of C-channels)	No	Yes
Number of digital links supported by one V5 interface	1	1–16

information flows called C-paths. A C-path carries messages of only one class and generally there is only one C-path per message type; but some message classes can be assigned to more than one C-path, to handle the traffic load. C-paths that carry the same class of messages must be in separate C-channels. C-paths and C-channels are used to manage (signaling) traffic capacity, not to identify message classes: the message class is indicated by the value of an IE carried in each message (see next sections).

ISDN D-channels can carry some user-data traffic in addition to signaling, and V5 standards distinguish three classes of D-channel data, based on the value of the SAPI parameter (Section 10.2):

1. ISDN Ds-data: D-channel data with SAPI other than 16 and 32–62. This type of data is used for call-control messages.
2. ISDN p-data: D-channel data with SAPI = 16 (X.25 packet data).
3. ISDN f-data: D-channel data with SAPI = 32–62 (Frame Relay data).

X.25 and Frame Relay are forms of packet communication (Chapter 20), which are not discussed in this book [4,5]. Each type of ISDN D-channel messages is assigned to a different C-path and each type can be carried by more than one C-path. Messages to/from the same line must stay in the same C-path (and thus in the same C-channel).

C-paths are assigned to C-channels and C-channels to time slots, semipermanently by provisioning. Bearer channels can be assigned to TSs semipermanently by provisioning or dynamically by signaling, depending on the protocol.

6.4 THE V5.1 STANDARD

V5.1 [6,7] supports only one digital link (E1) per interface and only fixed assignment of time slots to line interfaces, with no concentration. Time slots are assigned to user interfaces (as voice channels for analog lines and ISDN B-channels) and to C-channels by provisioning, with no signaling involvement.

Two protocols are used for signaling and interface management functions:

1. PSTN Signaling Protocol, for call control of analog lines.
2. Control Protocol, for status management of subscriber lines (analog and ISDN) and of the whole V5.1 interface.

Call control for ISDN lines is done via D-channel messages (Ds-data), which are passed semitransparently through the V5.1 interface.

The V5.1 protocol stack is shown in Fig. 6.4-1.

C-channels can be provisioned to carry the following C-paths: PSTN signaling messages, Control Protocol messages, ISDN Ds-data, ISDN p-data, and ISDN f-data. Up to three C-channels per V5.1 interface (i.e., per E1) are supported if

ISDN Application	PSTN Application	Control Application	Application
(null)	PSTN Protocol	Control Protocol	Network Layer (L3)
AN-FR	LAPV5-DL		Data Link Sublayers (L2)
LAPV5-EF			
C-Channel			Physical Layer (L1)

Figure 6.4-1. V5.1 protocol stack.

ISDN lines are present, and up to two if only PSTN lines are present. C-channels can be assigned to TS 16, 15, and 31, in that order, and are called C-channel 1, 2, and 3, respectively. The Control C-path must be in TS 16. The other C-paths can be in any C-channel. Time slots not assigned to C-channels can be provisioned for user interfaces.

6.4.1 V5.1 Data Link Layer (L2) Protocol

The data link layer protocol used by V5.1 is called *LAPV5* and is based on DSS1 LAPD (Section 10.2). LAPV5 is subdivided into three sublayers: *LAPV5-EF* (Envelope Function sublayer), *LAPV5-DL* (Data Link sublayer), and *AN-FR* (Frame Relay sublayer). Figure 6.4-1 shows the relationships between the sublayers.

LAPV5-EF Sublayer. This sublayer provides a common frame envelope for ISDN and other types of messages. One of its main functions is to identify the ISDN user port to which an ISDN message is related. For non-ISDN messages it identifies the protocol type but not the individual user port, which is identified by the L3addr IE of the Layer 3 protocols.

The format of LAPV5-EF frames is shown in Fig. 6.4-2. The class of messages contained in the Information fields is indicated by the values of the EFaddr

Bits 8	7	6	5	4	3	2	1	Octets
Flag 0	1	1	1	1	1	1	0	1
EFaddr						0	EA 0	2
EFaddr							EA 1	3
Information								4
…..								…
Information								n-3
FCS								n-2
FCS								n-1
Flag 0	1	1	1	1	1	1	0	n

EA: Address Extension Bit FCS: Frame Check Sequence

Figure 6.4-2. Frame format for the Envelope Function Sublayer. (From Rec. G.964. Reproduced with the kind permission of ITU.)

information element (IE). EFaddr values in the range 0–8175 are used for D-channel messages, with each value identifying a specific ISDN user port. Higher values indicate applications other than ISDN, with one value per application. The defined values are 8176 for PSTN Signaling messages and 8177 for Control Protocol messages.

LAPV5-DL Sublayer. This sublayer manages data link communication between L3 peer entities in the access system and the LE. It provides functions such as frame acknowledgments and repetition. The link access procedures of LAPV5-DL are essentially the same as the DSS1 LAPD procedures (Section 10.2). The full scope of LAPD is not utilized, however, due to the limited functionality required by V5.1. Specifically, LAPV5-DL uses:

- A subset of the standard Q.921 primitives
- A subset of the frame types (Table 6.4-1)
- Only multiframe acknowledged message transfers
- No TEI and related messages and procedures, since the connections for Control and PSTN signaling are fixed

The format for LAPV5-DL frames is shown in Fig. 6.4-3. Two types of frames are used: Type A, which has no Information fields (only the first four octets are present), and Type B, which has Information fields.

TABLE 6.4-1 Frame Types Used in the Control Fields of LAPV5-DL

Frame Type	Name
DM	Disconnected Mode
I	Information
REJ	Reject
RNR	Receive Not Ready
RR	Receive Ready
SABME	Set Automatic Balanced Mode Extended
UA	Unnumbered Acknowledgment

Bits

8	7	6	5	4	3	2	1	Octets
V5DLaddr						C/R	EA 0	1
V5DLaddr (lower)							EA 1	2
Control								3
Control								4
Information								5
...								...
Information								n

C/R: Command/Response EA: Address Extension Bit

Figure 6.4-3. Frame format for the Data Link Sublayer.

LAPV5-DL does not use the SAPI and TEI information elements of LAPD (Fig. 10.2-2) and replaces them with V5DLaddr. V5Dladdr identifies the higher layer application and has the same value as EFaddr for non-ISDN messages in the LAPV5-EF sublayer, namely, 8176 for PSTN Signaling and 8177 for Control Protocol. The C/R bit and the Control fields are as defined in Section 10.2.4.

AN-FR Sublayer. This sublayer handles the quasitransparent transport of ISDN D-channel signaling messages through the V5.1 interface. A Mapping Function (MF) encapsulates LAPD frames coming from the D-channel of ISDN lines into LAPV5-EF frames. The encapsulation consists of stripping flag and FCS, adding EFaddr, checking and recalculating the FCS, and adding the flag again. The opposite is done in the other direction. Through this mechanism, the LE communicates directly with the ISDN terminal equipment (TE), with no involvement by the V5.1 control logic, except for the simple mapping function.

6.4.2 V5.1 Layer 3 (L3) Protocols

Layer 3 consists of two protocols, the PSTN Signaling Protocol and the Control Protocol, which can be viewed as two aspects of the same protocol that handle different functions.

The general format of L3 messages is shown in Fig. 6.4-4. The first three IEs (Protocol Discriminator, L3 address, Message Type) are mandatory and present in all messages; message-dependent IEs are inserted after them, as needed.

The Layer 3 address IE (L3addr) identifies the analog subscriber line (15 bits) or the ISDN port (13 bits) to which the message is related. The PSTN Signaling Protocol uses only the 15-bit format. The Control Protocol uses both formats, since it controls both PSTN and ISDN lines. Control Protocol messages addressed to the common control function of the interface, and not related to individual lines, use the 13-bit format and the value 8177.

The list of messages is shown in Table 6.4-2.

Bits

8	7	6	5	4	3	2	1	Octets
Protocol Discriminator								1
0	1	0	0	1	0	0	0	
L3 Address						0*	1/0*	2
L3 Address (lower)							1*	3
0	Message Type							4
Information Elements								5
...								...
								n

* When L3addr has a 13-bit format, bits 1 and 2 of octet 2 are 0, 0, and bit 1 of octet 3 is 1. When L3addr has a 15-bit format, it uses all the bits in octets 2 and 3 except bit 1 of octet 2, which is set to 1.

Figure 6.4-4. Message format for the V5.1 L3 Signaling protocols. (From Rec. G.764. Reproduced with the kind permission of ITU.)

TABLE 6.4-2 Layer 3 Protocol Messages for V5.1

Protocol	Message Type	AS ↔ LE
PSTN	Disconnect	← →
	Disconnect Complete	← →
	Establish	← →
	Establish Acknowledgment	← →
	Protocol Parameter	←
	Signal	← →
	Signal Acknowledgment	← →
	Status	→
	Status Enquiry	←
Control	Common Control	← →
	Common Control Acknowledgment	← →
	Port Control	← →
	Port Control Acknowledgment	← →

PSTN Signaling Protocol. The main functions of this protocol are:

- Transmitting the electrical status of the analog line to the LE
- Instructing the AS to modify the electrical configuration of the AS toward the analog line
- Activating the TS (voice channel) associated with the subscriber line when a request for connection is received, and deactivating it when the connection is released (TS assignment is by provisioning)

The first two functions constitute a stimulus protocol for analog line supervision. To support the vast number of subscriber loop signaling versions used in different countries, V5.1 defines a large set of primitives for communication between the line control logic and the V5.1 signaling logic in the LE [6,7], as well as a large set of message-dependent IEs. The most significant IEs are shown in Table 6.4-3.

The basic signaling sequences for an AS-originated and a LE-originated call are shown in Figs. 6.4-5 and 6.4-6, respectively. The deactivation of the TS associated with the line is controlled by the LE. If the subscriber on the AS side hangs up first, the LE is informed by a Signal message.

Control Protocol. This protocol has two types of messages: Port Control messages used to manage the status of line interfaces, and Common Control messages used to manage the whole V5.1 interface. Examples of message-dependent IEs for the Control Protocol are shown in Table 6.4-4.

TABLE 6.4-3 PSTN Protocol Information Elements

IE ↓ Message →	Disconnect	Disconnect Complete	Establish	Establish Acknowledgment	Protocol Parameter	Signal	Signal Acknowledgment	Status	Status Enquiry[a]
Cadenced-ringing			C ←			C ←			
Enable-metering						C ←			
Line-information			C →						
Metering-report						C →			
Pulsed-signal			C ←	C ← →		C ← →			
Recognition-time					C ←				
Sequence-number					M ←	M ← →	M ← →		
State								M →	
Steady-signal	O ← →	O ←	C ← →	C ← →		C ← →			

[a]Status Enquiry does not have any IEs besides the mandatory first three.
C, Conditional (mandatory under certain conditions); M, Mandatory; O, Optional; →, AS to LE; ←, LE to AS.

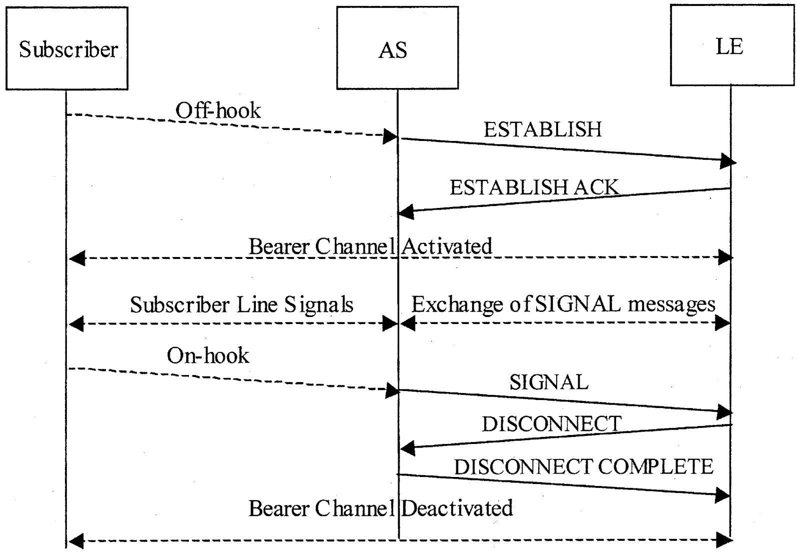


Figure 6.4-5. V5.1 PSTN signaling: AS-originated connection and release.

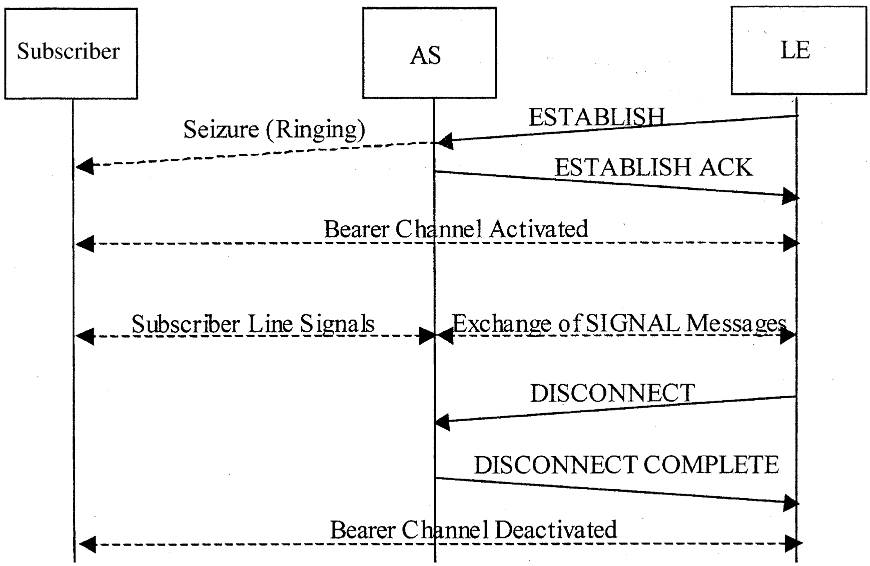


Figure 6.4-6. V5.1 PSTN signaling: LE-originated connection and release.

TABLE 6.4-4 Control Protocol Information Elements

IE ↓ Message →	Common Control	Common Control Acknowledgment	Port Control	Port Control Acknowledgment
Control-function-element			M ← →	M ← →
Control-function-ID	M ← →	M ← →		
Performance-grading			C →	

C, Conditional (mandatory under certain conditions); M, Mandatory; →, AS to LE; ←, LE to AS.

6.5 THE V5.2 STANDARD

6.5.1 Overview

The V5.2 interface is backward compatible with V5.1 [8,9] and introduces the following enhancements:

- Up to sixteen E1digital links per V5.2 interface
- Statistical multiplexing of voice channels (concentration)
- ISDN H-channels (multi-time-slot ISDN user channels)
- ISDN Primary Rate Interface (30B + D)

To support the new functionality, five Layer 3 protocols are used for signaling and interface management functions, the first two common with V5.1:

1. PSTN Signaling Protocol
2. Control Protocol
3. Link Control Protocol
4. Protection Protocol
5. Bearer Channel Control (BCC) Protocol

As in V5.1, signaling for call control of ISDN lines is carried in D-channel messages (Ds-data), passed semitransparently through the V5.2 interface.

The V5.2 protocol stack is shown in Fig. 6.5-1.

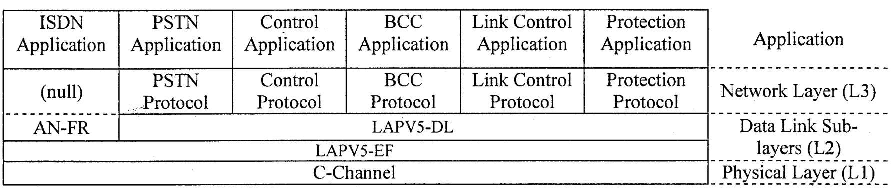


Figure 6.5-1. V5.2 protocol stack.

C-Channels. C-channels can be provisioned to carry the following C-paths: PSTN Signaling, Control, Link Control, Protection, BCC, ISDN Ds-data, ISDN p-data, and ISDN f-data.

C-channels are assigned to TS 16, 15, and 31 in that order. After all three TSs in a link are assigned to C-channels, the same-numbered TSs may be assigned to additional C-channels in other links.

One C-channel (assigned to TS 16) is always provisioned with at least the Control, Link Control, Protection, and BCC C-paths, thus serving all the digital links of a V5.2 interface. The other C-paths can be assigned to any C-channel.

Bearer Channels. These channels are assigned to TSs by the LE using the Bearer Channel Connection (BCC) Protocol. Three types of time-slot assignments are supported:

1. Dynamic assignment. This method allows concentration through statistical multiplexing.
2. Fixed assignment. This method (no concentration) can be used to avoid blocking. V5.2 recommendations use the term *preconnected bearer channel* to indicate this type of TS assignment.
3. Semipermanent (leased line) connections.

Time slots other than 16, 15, and 31 can be assigned without restrictions to analog lines or to ISDN B- and H-channels. Time Slots 15, 16, and 31 can also be assigned to user interfaces when not assigned to C-channels.

Protection Switching. This function is necessary in multilink configurations because some C-channels (like the basic C-channel) provide essential control communication for the whole V5 interface, so that the failure of one link may affect parts that are working properly. Two *protection groups* are configurable:

1. Protection Group 1—mandatory
2. Protection Group 2—optional

Protection Group 1 protects the basic C-channel. If only this protection group is provisioned, C-paths for PSTN and D-channel signaling are not protected. TS 16, assigned to the basic C-channel in one of the digital links (called the primary link), is the active C-channel. TS 16 in another link (called the secondary link) is the standby C-channel. The standby C-channel takes over as active when the primary link fails.

Protection Group 2 protects the other C-channels, which carry PSTN and ISDN signaling. Within this group, enough standby C-channels must be provisioned to backup all the C-channels in any single digital link.

Switching from active to standby C-channels is controlled by the Protection Protocol (described in Section 6.5.3).

No protection is supported for voice channels.

6.5.2 V5.2 Data Link Layer (L2) Protocol

The Envelope Function and the Data Link sublayer protocols are the same as for V5.1, except that three additional values are defined for the EFaddr and V5DLaddr, to accommodate the additional L3 protocols: 8178 for BCC, 8179 for Protection, and 8180 for Link Control.

6.5.3 V5.2 Layer 3 (L3) Protocols

Layer 3 consists of five protocols, all essentially variations of the same protocol. The basic message format is the same as for V5.1 messages. The first three IEs are always present, followed by message-dependent IEs (Fig. 6.4-4). The main difference from V5.1 is in the Layer 3 address, which in V5.1 is called L3addr. In V5.2 the Layer 3 address occupies the same bit positions (two octets) as L3addr of V5.1 but has different names and formats depending on the L3 protocol, as shown in Table 6.5-1. The bit-level formats of L3addr for the PSTN Signaling and the Control protocols are the same as for V5.1; the others are slightly different and, for them, the reader is referred to [9].

The main L3 protocol messages for V5.2 are shown in Table 6.5-2.

PSTN Signaling Protocol. This protocol is used for call control of analog lines and is the same as for V5.1.

Control Protocol. This protocol performs the same function and is basically the same as the corresponding protocol for V5.1, with a few modifications to support ISDN Primary Rate Access:

- Additional values for the Control Function ID information element.
- Additional values for the Control Function Element IE.
- Primary rate ports are permanently activated so there is no activate/deactivate procedure, resulting in an accelerated alignment procedure.

TABLE 6.5-1 Layer 3 Address Information Elements for V5.2

Protocol	Information Element Name	Entity Identified
PSTN Protocol	L3addr	PSTN user port (analog line)
Control Protocol	L3addr	ISDN/PSTN user port or common control function
Link Control Protocol	L3addr	E1 link
BCC Protocol	BCC Reference Number	BCC protocol process
Protection Protocol	Logical C-Channel Identification	Logical C-channel

TABLE 6.5-2 Layer 3 Protocol Messages for V5.2

Protocol Subset	Message Types	AS ↔ LE
PSTN Protocol	Disconnect	← →
	Disconnect Complete	← →
	Establish	← →
	Establish Acknowledgment	← →
	Protocol Parameter	←
	Signal	← →
	Signal Acknowledgment	← →
	Status	→
	Status Enquiry	←
Control Protocol	Common Control	← →
	Common Control Acknowledgment	← →
	Port Control	← →
	Port Control Acknowledgment	← →
Protection Protocol	OS Switch-Over Command	←
	Protocol Error	→
	Switchover Acknowledgment	→
	Switchover Command	←
	Switchover Request	→
	Switchover Reject	← →
BCC Protocol	Allocation	←
	Allocation Complete	→
	An Fault	→
	An Fault Acknowledgment	←
	Audit	←
	Audit Complete	→
	Deallocation	←
	Deallocation Complete	→
Link Control Protocol	Link Control	← →
	Link Control Acknowledgment	← →

Link Control Protocol. This protocol is used to manage multiple digital links in an interface and supports functions such as:

- Link status (in service, out of service, awaiting link control acknowledgment)
- Link blocking and unblocking
- Checking link continuity

The protocol is symmetrical; that is, the procedures work the same way in either direction. Messages have only one additional IE (Link Control function) after the first three mandatory elements of the basic message format.

Protection Protocol. The function of this protocol is to switch a C-channel from active to a standby and vice versa. The switch can be triggered by link failure or, for Protection Group 2 only, by operator command. The choice of the standby C-channel that takes over as active is controlled by the LE, although the operator

TABLE 6.5-3 Protection Protocol Information Elements

IE ↓ Message →	OS-Switch-over Complete	Protocol Error	Switchover Acknowledgment	Switch-over Complete	Switch-over Request	Switch-over Reject
Sequence number	M	M		M	M	M
Physical C-channel identification	M		M	M	M	M

M, Mandatory. Each IE applies in the same direction(s) as the message.

TABLE 6.5-4 BCC Protocol Information Elements

IE ↓ Message →	Allocation	Deallocation	Audit	Audit Complete	An Fault
User port identification	M	M	O	O	O
ISDN port TS identification	C	C	O	O	O
V5 time slot identification	C	C	O	O	O
Multislot map	C	C			

C, Conditional (mandatory under certain conditions); M, Mandatory; O, Optional. Each IE applies in the same direction(s) as the message.

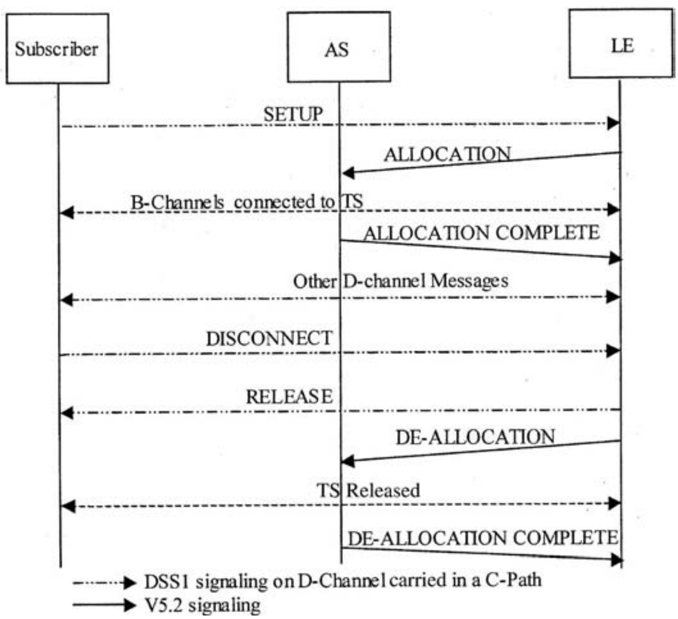


Figure 6.5-2. V5.2 ISDN call: AS-originated connection and release.

can specify a preference. Switchover does not change the configuration of C-paths in a given C-channel. Examples of IEs in Protection Protocol messages are shown in Table 6.5-3 (the first three mandatory IEs are not shown).

BCC Protocol. This protocol is used for the assignment of time slots to user interfaces. As explained in Section 6.5.1, the assignment can be dynamic, fixed, and semipermanent.

The main IEs are shown in Table 6.5-4 (the first three mandatory IEs are not shown). In Table 6.5-4 the messages not shown do not have any of the IEs listed

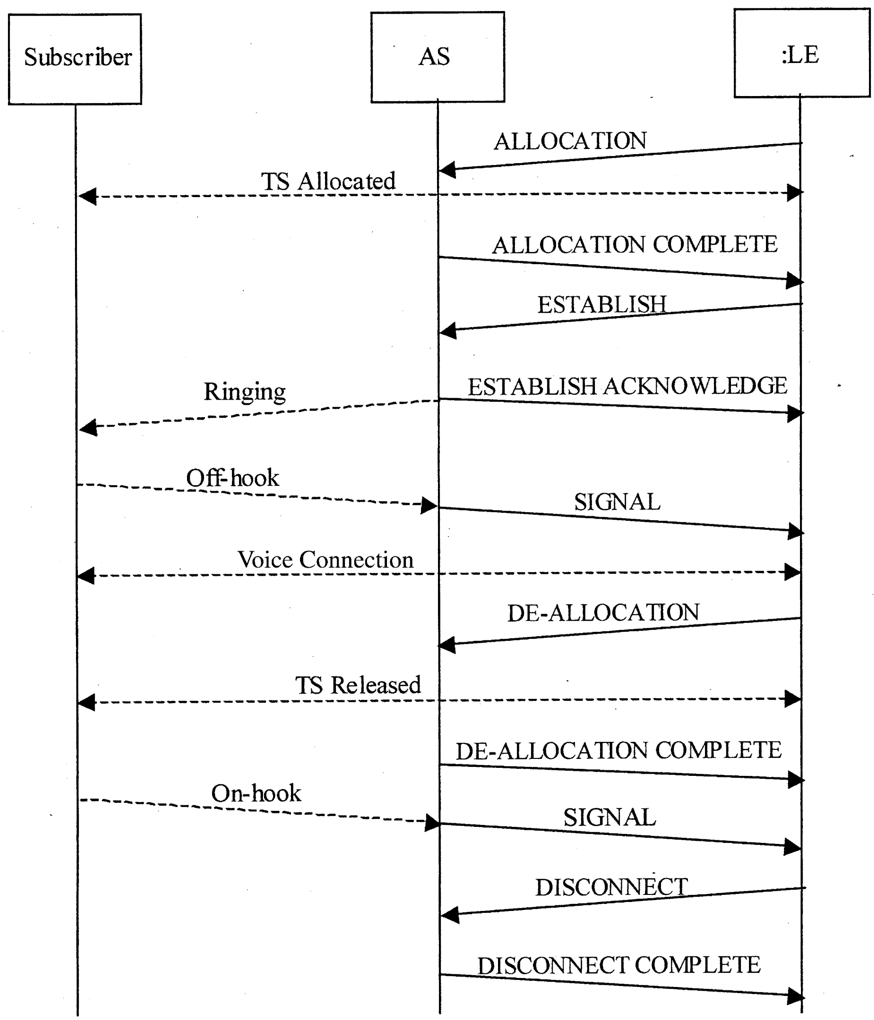


Figure 6.5-3. V5.2 PSTN call: LE-originated connection and release.

in the table. The User Port Identification IE identifies the analog line or ISDN port to which the message is related: for PSTN lines it uses the same 15-bit values as L3addr in the PSTN L3 protocol, and for ISDN lines the same 13-bit values as the EFaddr in the LAPV5-EF sublayer. The ISDN Port TS Identification IE identifies the B-channel that needs to be connected to a TS (e.g., channel B1 or B2 of a basic rate interface or channel 1 through 31 of a primary rate interface). The V5 Time Slot Identification IE identifies the digital link and the TS in that link that is assigned to the user interface. The Multislot Map IE is used when allocating blocks of TSs to H-channels ($n \times 64$ kb/s).

Call Sequences. Examples of call sequences are shown in Figs. 6.5-2 and 6.5-3. In the case of ISDN calls, two allocation processes, both controlled by the LE, take place: B-channels are assigned to the call and TSs to B-channels. The first allocation (B- or H-channel assignment, e.g., B1 or B2) is part of standard ISDN call-setup procedures (Section 10.4.2); the second is specific to V5.2. Both allocations are communicated by the LE to the AS in a BCC message, but the first is also sent from the LE directly to the ISDN customer-premises equipment in a D-channel message. B-channel identification is contained in the Channel Identification IE, which for LE originations is carried in the Setup message, and for AS originations is carried in the response to the Setup message (e.g., Alerting or Call Proceeding).

6.6 ACRONYMS

AN	Access network
AN-FR	Access network frame relay
AS	Access system
BA	Basic access
BCC	Bearer Channel Control
BRA	Basic rate access
C	Conditional
CCC	Clear channel connection
C-channel	Communication channel
CSC	Common Signaling Channel
DS0	Digital signal level 0
DS1	Digital signal level 1
DSS1	Digital Subscriber Signaling System No. 1
DTMF	Dual-tone multi frequency
EOC	Embedded Operations Channel
ESF	Extended superframe
ET	Exchange termination
ETSI	European Telecommunications Standards Institute
FCS	Frame check sequence
ID	Identity
IDLC	Integrated digital loop carrier

IDT	Integrated digital terminal
IE	Information element
ISDN	Integrated Services Digital Network
ITU	International Telecommunication Union
ITU-T	Telecommunication Standardization Sector of ITU
L1	Layer 1
L2	Layer 2
L3	Layer 3
LAPD	Link Access Protocol D-Channel
LAPV5-DL	Link Access Protocol V5 Data Link
LAPV5-EF	Link Access Protocol Envelope Function
LC	Loop closure
LDS	Local digital switch
LE	Local exchange
M	Mandatory
MF	Mapping Function
NA	Not applicable
O	Optional
OAM	Operations, administration, and maintenance
OS	Operations system
PRA	Primary rate access
PSTN	Public switched telephone network
RDT	Remote digital terminal
SAPI	Service access point indicator
SDH	Synchronous digital hierarchy
SONET	Synchronous optical network
STM	Synchronous transport module
TDM	Time-division multiplexing
TEI	Terminal endpoint identifier
TMC	Timeslot Management Channel
TS	Time slot

6.7 REFERENCES

1. *Integrated Digital Loop Carrier System Generic Requirements, Objectives, and Interfaces*, GR-303-Core Issue 4, Telcordia Technologies, Dec. 2000.
2. *ISDN User–Network Interface—Data Link Layer Specification*, Rec. Q.921, ITU-T, Geneva, Sept. 1997.
3. *Digital Exchange Interfaces for Subscriber Access*, Rec. Q.512, ITU-T, Geneva, Feb. 1995.
4. *Interface Between Data Terminal Equipment (DTE) and Data Circuit-Terminating Equipment (DCE) for Terminals Operating in the Packet Mode and Connected to Public Data Networks by Dedicated Circuit*, Rec. X.25, ITU-T, Geneva, Oct. 1996.
5. *Framework for Frame Mode Bearer Services*, Rec. I.122, ITU-T, Geneva, Mar. 1993.

6. *V Interfaces at the Digital Local Exchange (LE); V5.1 Interface for the Support of Access Network (AN); Part 1: V5.1 Interface Specification*, EN 300 324-1 V2.1.1, ETSI, Sophia Antipolis, France, Apr. 2000.
7. *V-Interfaces at the Digital Local Exchange (LE)—V5.1 Interface (Based on 2048 kbit/s) for the Support of Access Network (AN)*, Rec. G.964, ITU-T, Geneva, Mar. 2001.
8. *V Interfaces at the Digital Local Exchange (LE); V5.2 Interface for the Support of Access Network (AN); Part 1: V5.2 Interface Specification*, EN 300 347-1 V2.2.2, ETSI, Sophia Antipolis, France, Dec. 1999.
9. *V-Interfaces at the Digital Local Exchange (LE)—V5.2 Interface (Based on 2048 kbit/s) for the Support of Access Network (AN)*, Rec. G.965, ITU-T, Geneva, Mar. 2001.

7

INTRODUCTION TO SIGNALING SYSTEM NO. 7

CCITT (now ITU-T) began the specification of the second-generation common-channel signaling system, known originally as CCITT No. 7, in the mid-1970s. The system was developed initially for telephony call control, and the first recommendations were published in the CCITT *Yellow Books* of 1981. Since then the applications of SS7 signaling have expanded and now include call control of the Integrated Services Digital Network (ISDN), as well as operations in the network that are not related to individual trunks. Most of that work was published in the CCITT *Blue Books* of 1989. More recent additions have also been issued and all recommendations are now published individually by the ITU-T. The suite of protocols is called Signaling System No. 7 (SS7).

Signaling System No. 6 and CCIS (Chapter 5) were used only in the international network and in the AT&T long-distance network, respectively. The objective for SS7 was to develop a signaling system that could be used worldwide. That objective, however, has not been met completely. Several national versions have been defined, notably the version specified by ANSI and Bellcore (now Telcordia) for the United States, and British Telecom's National User Part (NUP), for the United Kingdom.

In what follows we speak of SS7 when discussing the general aspects of the system that essentially apply to all versions, we refer to the ITU-T-defined system as Signaling System No. 7 (SS No. 7), and we refer to the North American version as ANSI No. 7.

SS7 Signaling Links. The signaling links are carried by the digital transmission channels of digital (PCM) trunks. In the United States those signaling links generally transmit at 56 kb/s. In most other countries the transmission is at 64 kb/s [1].

Signaling in Telecommunication Networks, Second Edition, by John G. van Bosse and Fabrizio U. Devetak

7.1 SS7 STRUCTURE

SS7 is defined in terms of messages and functions; hardware architecture issues are left to equipment manufacturers. The signaling system is divided into a number of *protocols* (or *parts*), each of which handles a group of related functions. The interfaces between those parts were defined in the early stages of the specification work. The various parts could then be specified simultaneously and independently of each other. That has made the overall specification effort more manageable.

7.1.1 SS7 Hierarchy

The parts of SS7 are organized in a four-level hierarchy [1–4]. We say that a higher-level part is a *user* of the services provided by a lower-level part. This arrangement is similar to the seven-layer structure of the Open Systems Interconnection (OSI) used for data communication protocols (Chapter 20). Several efforts have been made to align the SS7 levels and OSI layers. The efforts, however, have been only partially successful. OSI layers 1 and 2 correspond to SS7 levels 1 and 2, but things start to differ in the higher parts of the hierarchies. In what follows the four-level SS7 hierarchy will be used. It works well for trunk-related applications, and less well for other applications.

7.1.2 SS7 Protocols

The protocols of SS7 and their levels are shown in Fig. 7.1-1.

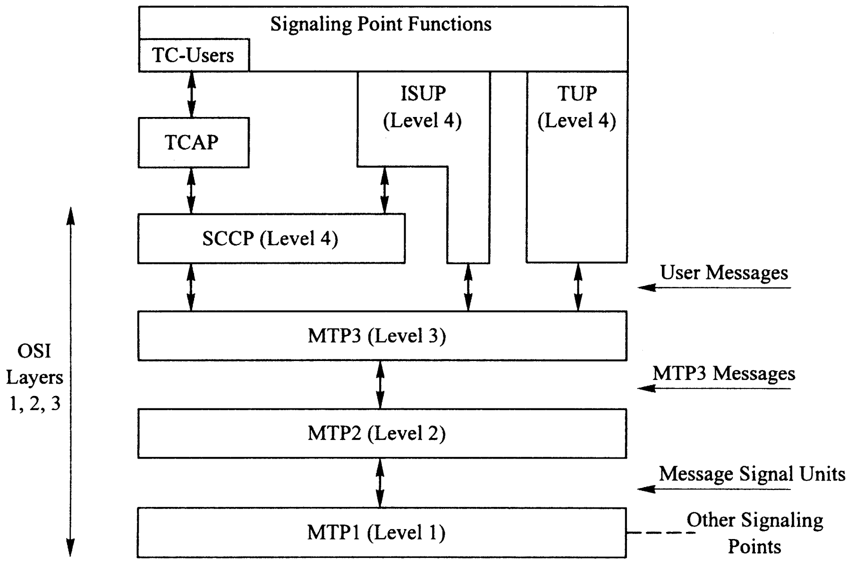


Figure 7.1-1. Structure of SS7. (From Rec. Q.700. Courtesy of ITU-T.)

Message Transfer Part (MTP). This protocol provides message transfer services for its *users*. It is divided into three parts—denoted as MTP1, MTP2, and MTP3—that occupy levels l, 2, and 3 of the SS7 hierarchy.

A MTP-user passes its outgoing messages to—and receives its incoming messages from—the MTP3 at its signaling point. A signaling point, has one, MTP3.

A combination of a MTP2 and a MTP1 represents a signaling link at a signaling point (a signaling link between two signaling points consists of a MTP1/MTP2 combination at each point). A signaling point that terminates *n* signaling links has *n* of these combinations.

Telephone User Part (TUP). TUP (a MTP-user) is a protocol for telephony call control and for trunk maintenance. It is very similar to SS6 but includes a number of additional features.

Integrated Services User Part (ISUP). ISUP (another MTP-user) is a protocol for call-control and trunk-maintenance procedures in both the telephone network and the ISDN.

TUP and ISUP messages are trunk-related: they contain information concerning a particular trunk with TUP or ISUP signaling and are sent by the exchange at one end of the trunk to the exchange at the other end.

Signaling Connection Control Part (SCCP). This protocol (a MTP-user) provides functions for the transfer of messages that are not trunk-related. Its users are ISUP and TCAP. SCCP does not fit neatly into the four-level hierarchy, because it is at the same level as ISUP.

The combination of MTP and SCCP corresponds to OSI layers 1, 2, and 3 and is known as the *network services* part of SS7.

Transaction Capabilities Application Part (TCAP). Transactions are operations that are not related to individual trunks and involve two signaling points (Chapter 16). The TCAP protocol provides a standard interface to *TC-users* (functions at a signaling point that ate involved in transactions of various kinds). In turn, TCAP is a user of SCCP.

7.1.3 Messages and Message Transfer

In Fig. 7.1-1 we distinguish messages of several types in a signaling point. *User messages* are messages between a level-4 protocol and the MTP3 and are named after the level-4 protocol: we speak of TUP, ISUP, and SCCP messages. *MTP3 messages* are messages between the MTP3 and a MTP2. *Message signal units* (MSUs) are messages between the MTP2 and MTP1 of a signaling link at a signaling point, and between the MTP1s at both ends of a signaling link. A MSU contains a message originated by MTP3, or by a MTP3-user.

User messages are transferred between two "peer" level-4 protocols at two signaling points. Figure 7.1-2 illustrates the transfer of a TUP message from TUP-A

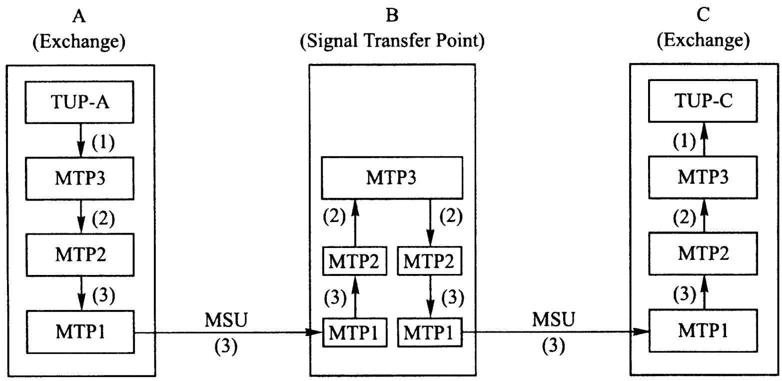


Figure 7.1-2. Transfer of a TUP message. (1), TUP message; (2), MTP3 message; (3), message signal unit.

to TUP-C (at signaling points A and C, respectively) that is routed via signal transfer point B.

At signaling point A, TUP-A passes the TUP message *downward* to its MTP3, which expands it into a MTP3 message and passes it to the MTP2 of the signaling link to B. MTP2 expands the MTP3 message into a message signal unit (MSU) and passes it to its MTP1.

The MSU traverses the signaling link and arrives at the MTP1 of signaling point B, where MTP2 extracts the MTP3 message and passes it to its MTP3. MTP3 transfers the MTP3 message to the MTP2 of the signaling link to C.

The second leg of the message transfer is similar: the message is passed and expanded downward in signaling point B, traverses the signaling link between B and C (as a MSU), and is passed upward in signaling point C. It finally arrives as a TUP message at TUP-C.

7.2 IDENTIFICATION OF SIGNALING POINTS AND TRUNKS

Signaling system No. 7 identifies signaling points (exchanges, service control points, etc.) and trunks with parameters that have a nationwide scope.

7.2.1 Point Codes

Each SS7 signaling point in a network is identified by a *point code* (PC). Most signaling points have only one (national) point code. However, an *international switching center* (ISC) is identified in the network of its country by a national PC and in the international network by an international PC.

7.2.2 ANSI No. 7 Point Codes

The format of ANSI No.7 point codes (used in the United States only) are shown in Fig. 7.2-1(*a*). The PC has three eight-bit fields that contain the parameters N, C, and

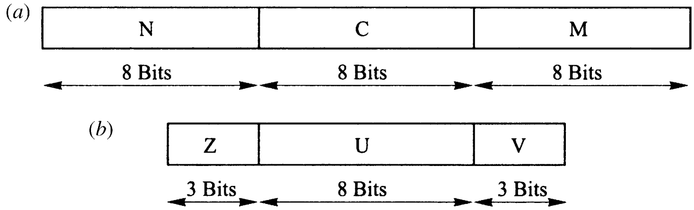


Figure 7.2-1. Point code formats: (*a*) ANSI No. 7 and (*b*) CCITT No. 7. (From Rec. Q.708. Courtesy of ITU-T.)

M [1]. Parameter N identifies the network of a particular telecom in the United States (Verizon, Sprint, AT&T, etc.). Parameter C represents a cluster of signaling points within a network. For example, a particular value of C may identify the exchanges with signaling links to a particular pair of STPs. Parameter M identifies a signaling point within a cluster. The ANSI No. 7 point codes are assigned by Telcordia.

7.2.3 International Point Codes

Point codes for the international network are assigned by ITU-T [5] and consist of parameters Z and V (3 bits each) and U (8 bits)—see Fig. 7.2-1(*b*).

Z identifies six major geographical *world zones.* In decimal representation:

2 Europe
3 North America
4 Mideast and most of Asia
5 Australia and part of Asia (Indonesia, Malaysia, Thailand, Guam, etc.)
6 Africa
7 Latin America

These world zones are different from the world zones represented by the initial digits in country codes (Section 1.2.3).

Parameter U identifies an area of a national telecommunication network within a world zone. Most countries have only one network that covers the entire country and are therefore represented by one value of U. For example, the combination Z = 2, U = 168 represents the national network of Bulgaria.

V identifies a particular international switching center in the network—or network area—specified by Z and U. In countries with one national network, up to seven ISCs can be identified.

Countries with more than one national network have several U codes. For example, the U.K. national networks operated by British Telecom and Mercury Telecommunications are identified by Z = 2, U = 068 and Z = 2, U = 072.

In the United States, values of U are assigned to particular areas of the long-distance networks operated by the various international carriers (AT&T, Verizon, Sprint, etc.). For example, a value of U identifies the East Coast area of AT&T's long-distance network, and V identifies an ISC in that area of the network.

ITU-T has allocated the range U = 020–059 in zone Z = 3 to the United States.

7.2.4 National Point Codes in Other Countries

Countries other than the United States use 14-bit national point codes. The code assignments are made by the individual national telecoms.

7.2.5 Identification of Trunks

A trunk group with TUP or ISUP signaling is identified uniquely in a national network (or in the international network) by the point codes of the exchanges that are interconnected by it.

The *circuit identification code* (CIC) identifies a trunk within a trunk group. The CIC field has a length of 12 bits and thus can identify trunks in groups of up to 4095 trunks.

7.3 SS7 SIGNAL UNITS AND PRIMITIVES

SS7 signal units (SUs) have different lengths but always occupy an integral number of octets (groupings of eight bits). A SU consists of an ordered set of fields with parameters. In the documentation of the early SS7 parts (MTP and TUP), the SUs are shown as in Fig. 7.3-1(*a*). Par.1, Par.2, ... denote the first, second, ... parameter fields of the SU. In the more recently defined parts of SS7, a SU is shown as a stack of octets—see Fig. 7.3-1(*b*). Bits in the octets are numbered from right to left, and bit 1 of octet 1 is the first bit sent out. For uniformity, this representation is used throughout this book.

The lengths of SS7 parameter fields are not limited to integral multiples of octets. For example, Par.2 and Par.3 in Fig. 7.3-1(*b*) have a length of 12 bits.

A working signaling link carries a continuous SU stream in each direction.

7.3.1 Signal Unit Types

In addition to the message signal unit, SS7 includes two other SU types [6]. The *link status signal units* (LSSUs) and *fill-in signal units* (FISUs) originate in the MTP2 at one end of the signaling link and are processed by the MTP2 at the other end (they are the SS7 equivalents of the "link" SUs described in Section 5.2). LSSUs are used for the control of a signaling link, and FISUs are sent when no MSUs or LSSUs are waiting to be sent out.

Figure 7.3-2 shows a MSU in some detail. It consists of a MTP3 message, surrounded by MTP2 data. Length indicator (LI) has a dual role. In the first place, it indicates the number of octets, measured from octet 4 through octet $(n-2)$. In addition, the value of LI implies the SU type. In MSUs, LI exceeds 2, LSSUs have LI = 1 or 2, and FISUs have LI = 0.

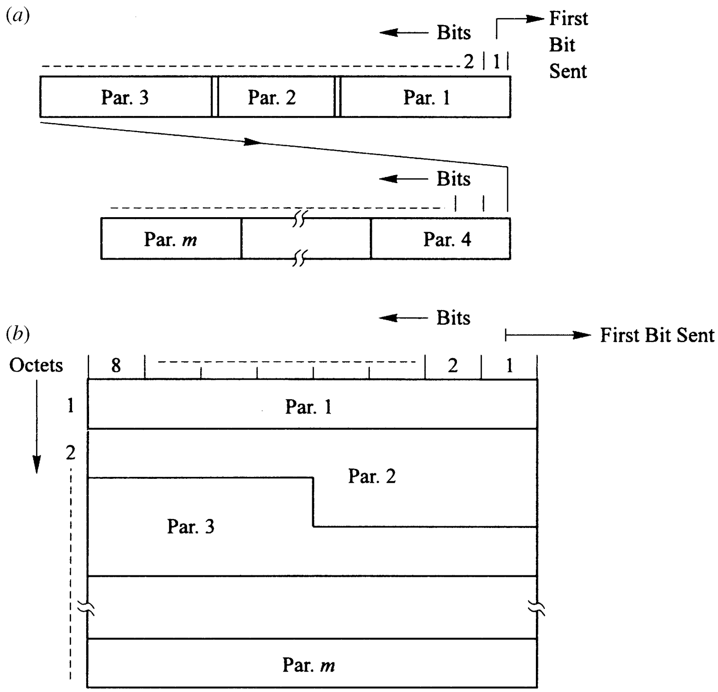


Figure 7.3-1. Representation of signal units: (*a*) in MTP and TUP documents and (*b*) in ISUP, SCCP, and TCAP documents.

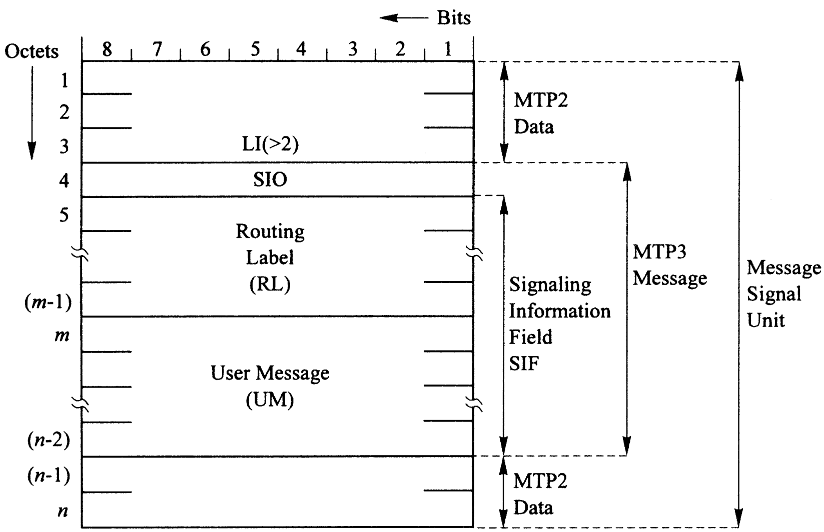


Figure 7.3-2. Message signal unit. (From Rec. Q.700. Courtesy of ITU-T.)

The MTP2 data are added by the MTP2 of the signaling point where the message originates and are processed and then removed by the MTP2 at the signaling point on the other end of the signaling link.

The *MTP3 message* consists of the *service information octet* (SIO) and the *signaling information field* (SIF). The maximum length of SIF is 272 octets.

SIO identifies the MTP-user (TUP, ISUP, or SCCP) that has originated the message and is used by the MTP3 in the destination signaling point to deliver the message to the "peer" user.

The SIF consists of the *routing label* (RL) and the *user message* (UM). The routing label contains parameters that are used by the MTP3s in the signaling points along the message path to route the MSU to its destination signaling point. The user message contains information for the MTP-user at the destination signaling point.

The user message and the information in SIO and RL are supplied by the originating MTP-user. The parameters in SIO and RL are interface parameters and are used by the MTP3s and by the MTP-user at the destination signaling point. The user message is passed transparently (not examined by MTP3).

7.3.2 Primitives

In the ITU-T model of SS7, the messages between protocols in a signaling point are passed in standardized interface elements called *primitives* [7]. There are four types of primitives: *requests* and *responses* pass information from a higher-level protocol to a lower-level protocol, and *indications* and *confirmations* pass information in the opposite direction.

Primitives between two protocols are named after the lower-level protocol. For example, the primitives for the message transfer between MTP3 and the MTP-users are known as *MTP-transfer* primitives.

ITU-T has defined the information to be included in each primitive. Since primitives pass information inside a signaling point only, their (software) implementation is left to the equipment manufacturers.

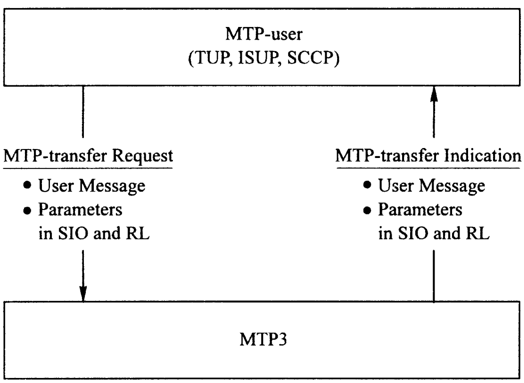


Figure 7.3-3. Message transfer primitives. (From Rec. Q.701. Courtesy of ITU-T.)

Figure 7.3-3 shows the message transfer primitives between MTP3 and a MTP-user at a signaling point. On outgoing messages, the user passes a MTP-transfer request that includes the user message and the parameters in SIO and RL. On incoming messages, MTP3 delivers these data to the MTP-user, in a MTP-transfer indication that contains the same information.

7.4 ACRONYMS

ANSI	American National Standards Institute
CB	Check bit
CCIS	Common-Channel Interoffice Signaling
CCITT	International Telegraph and Telephone Consultative Committee
CIC	Circuit identification code
F	Flag
FISU	Fill-in signal unit
ISC	International switching center
ISO	International Standards Organization
ISDN	Integrated Services Digital Network
ISUP	ISDN User Part
ITU	International Telecommunication Union
ITU-T	Telecommunications Standardization Sector of ITU
kb/s	Kilobits per second
LI	Length indicator
LSSU	Link status signal unit
MSU	Message Signal unit
MTP	Message Transfer Part
OSI	Open Systems Interconnection
PC	Point code
PCM	Pulse code modulation
RL	Routing label
SCCP	Signaling Connection Control Part
SIF	Signaling information field
SIO	Service information octet
SL	Signaling link
SP	Signaling point
SS6	Signaling System No. 6
SS7	Signaling System No. 7
STP	Signal transfer point
SU	Signal unit
TCAP	Transaction Capabilities Application Part
TUP	Telephone User Part
UM	User message

7.5 REFERENCES

1. A. R. Modaressi and R. A. Skoog, Signaling System No. 7: A Tutorial, *IEEE Commun. Mag.*, **28**, No. 7, July 1990.
2. R. Manterfield, *Common Channel Signalling*, Peter Peregrinus Ltd. London, 1991.
3. K. G. Fretten and C. G. Davies, CCITT Signalling System No. 7: Overview, *Br. Telecommun. Eng.*, **7**, Apr. 1988.
4. *Specifications of Signaling System No. 7*, Rec. Q.700, ITU-T, Geneva, 1993.
5. *Specifications of Signalling System No. 7*, Rec. Q.708, ITU-T, Geneva, 1999.
6. *Specifications of Signalling System No. 7*, Rec. Q.703, ITU-T, Geneva, 1996.
7. *Specifications of Signalling System No. 7*, Rec. Q.701, ITU-T, Geneva, 1993.

8

SS7 MESSAGE TRANSFER PART

The Message Transfer Part (MTP) of SS7 has two main functions. In the first place, it handles the transfer of MTP-user messages across the SS7 signaling network. In the second place, it includes functions to keep the message traffic flowing when failures occur in the signaling network.

8.1 INTRODUCTION TO MTP

8.1.1 Structure of MTP

MTP is divided into three parts, located at levels 1, 2, and 3 of the SS7 hierarchy. The main functions of these parts are outlined below [1–5].

MTP Level 1 (MTP1). MTP1 is the physical *signaling data link* (SDL), which consists of a pair of 64-kb/s digital transmission channels, and transports SS7 signal units between two signaling points. MTP1 is described in Section 8.2.

MTP Level 2 (MTP2). A *signaling link* (SL) between signaling points A and B consists of a SDL between the signaling points and MTP2 functions located at both signaling points (Fig. 8.1-1). The MTP2 functions relate to individual signaling links and include synchronization and the detection and correction of errors in *message signal units* (MSUs). MTP2 is discussed in Sections 8.3 through 8.6.

MTP Level 3. MTP3 is the interface between MTP and the MTP-users (level-4 protocols) at a signaling point. In addition to providing services for the transfer of user messages, MTP3 includes procedures to reroute messages when a failure occurs in the SS7 signaling network. MTP3 is discussed in Sections 8.7 through 8.9.

Signaling in Telecommunication Networks, Second Edition, by John G. van Bosse and Fabrizio U. Devetak

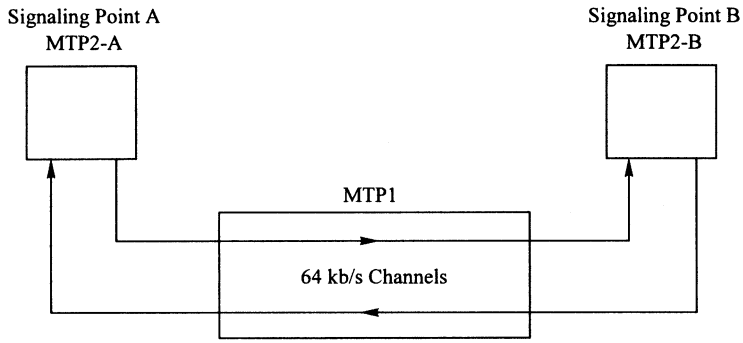


Figure 8.1-1. SS7 signalling link.

At a signaling point, there is a MTPl and a MTP2 for each signaling link and a single MTP3 (Fig. 8.1-2). The signaling links carry message signal units, *link status signal units* (LSSUs), and *fill-in signal units* (FISUs). The LSSUs and FISUs originate in the MTP2 at one end of a signaling link and terminate in the MTP2 at the other end.

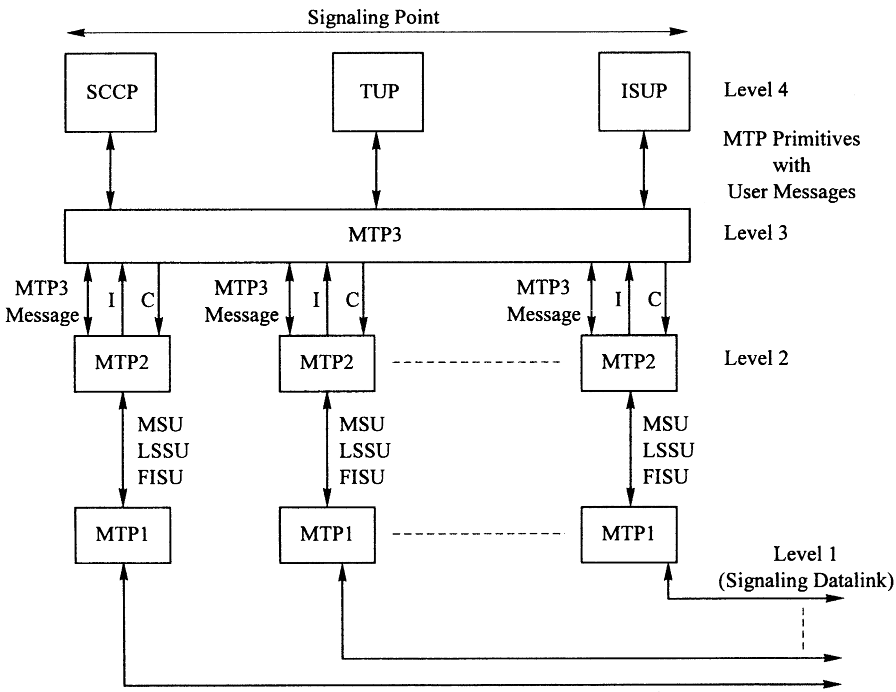


Figure 8.1-2. MTP structure.

8.1.2 Message Transfer

A MTP-user passes its outgoing message to MTP3, in a MTP-transfer primitive. MTP3 expands the user message into a MTP3 message, selects the outgoing signaling link, and passes the MTP3 message to the MPT2 of that link. The MTP2 expands the MTP3 message into a MSU, which is sent out.

A MTP2 extracts the MTP3 message from a received MSU and passes it to MTP3. MPT3 extracts the user message and passes it to the appropriate user.

8.2 MTP LEVEL 1

MTP1 defines the physical aspects of SS7 signaling data links, which are carried by digital 64-kb/s time-division channels (time slots) of digital transmission systems [6].

In today's telecommunication networks, the number of digital trunks exceeds the number of SS7 signaling links by about two orders of magnitude. It is not economical to install digital transmission systems that are dedicated to signaling data links. Instead, the digital transmission systems are shared by trunks and signaling data links.

Figure 8.2-1 shows a building that houses an exchange that has SS7 trunks and a signal transfer point. The digital transmission systems (a) that enter the building from other buildings of a telecommunication network can be first- or higher-order transmission systems. Some of these carry trunks only, and others carry trunks and signaling links. After demultiplexing the higher-order systems, a number of first-order multiplexes (b) are obtained. These multiplexes have 24 or 30 digital 64-kb/s channels (Section 1.5.2). The multiplexes are attached to a *digital cross-connect system.*

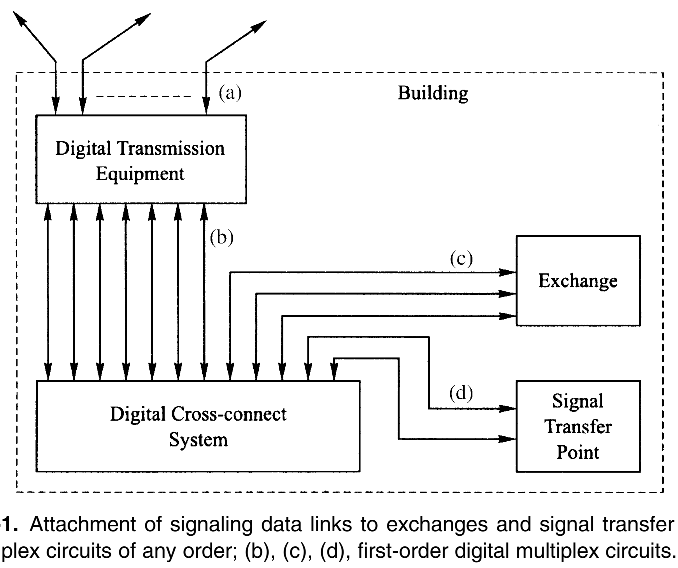


Figure 8.2-1. Attachment of signaling data links to exchanges and signal transfer points. (a), Digital multiplex circuits of any order; (b), (c), (d), first-order digital multiplex circuits.

Digital cross-connect systems are very similar to the digital switchblocks in exchanges (Section 1.7.2). They have a number of ports to which first-order digital multiplexes are attached, and the paths in cross-connect systems are bidirectional 64-kb/s paths that connect a channel (time slot) of one multiplex to a channel of another multiplex. However, unlike the paths in exchange switchblocks, these paths are semipermanent and are established and released under command of telecom personnel.

In Fig. 8.2-1, each channel in the multiplexes (b) corresponds to a channel in one of the digital transmission systems (a). The cross-connect system segregates the channels attached to the exchange and the channels attached to the signal transfer point. This is done by setting up paths such that the channels in the multiplexes (c) are associated with the digital trunks and SS7 signaling data links of the exchange, and the channels in multiplexes (d) carry the signaling data links of the signal transfer point.

Transfer Rates of Signaling Data Links. The channels of North American first-order digital multiplex transmission systems (T_1 systems) nominally transfer bits at a rate of 64 kb/s—see Fig. 1.5-4(*a*). However, there is a limitation on bit patterns: at least one bit in every time slot should be a "1" (pulse). This is a minor problem for time slots that carry the speech samples of PCM trunks and is solved by substituting the bit pattern 0000 0000 by the pattern 0000 0001. This causes a small—but acceptable—distortion in the decoded analog waveform.

The sequences of 1's and 0's on a SS7 signaling data link are unpredictable. Therefore, in the time slots associated with signaling data links, bit 8 is permanently set to "1." This reduces the transfer rate of SS7 signaling links carried by these transmission systems to $8000 \times 7 = 56$ kb/s.

The first-order digital multiplexed transmission systems defined by ETSI and used in most countries outside North America, use a form of transmission that can handle any sequence of 1's and 0's. SS7 signaling links on these transmission systems can therefore operate at 64 kb/s.

8.3 OVERVIEW OF MTP LEVEL 2

In a signaling point, each signaling data link is connected to a MTP2. The MTP2 functions are similar to those of the signaling terminals in SS6 signaling. The primary MTP2 responsibility is to transfer MSUs across the signaling link, including error detection and correction. MTP2 also monitors and controls the status of the link [5,7].

The reliability objectives for SS7 signaling links are as follows [8]:

1. The probability that errors in a received MSU are not detected should be less than one in 10^{10}.
2. Failures should not cause the loss of more than one MSU in 10^7.
3. Less than one in 10^{10} messages should be delivered out-of-sequence to the user parts.

In earlier CCS systems, error correction consisted of retransmitting a message signal unit that has incurred transmission errors. The retransmitted MSU arrived after MSUs that were sent later and did not incur errors (out-of-sequence delivery).

Since the *bit error rate* on typical signaling links is in the range of 10^{-4} to 10^{-6}, objective 3 dictates an error-correction procedure that does not cause out-of-sequence message delivery. This is accomplished as follows. When a MTP2 receives a mutilated message signal unit, say, MSU_m, it discards that MSU *and* all subsequently received MSUs, until it receives an error-free retransmitted copy of MSU_m. Conversely, when the distant MTP2 receives the negative acknowledgment of MSU_m, it retransmits that MSU *and* all subsequently sent MSUs.

SS7 error correction is discussed in Sections 8.4 and 8.5. By eliminating the major cause of out-of-service MSU delivery, SS7 allows simpler (and more drastic) procedures to cope with the receipt of an out-of-sequence MSU.

8.3.1 MTP2 Structure

The main parts of MTP2 are shown in Fig. 8.3-1. *Link control* (LC) controls the other functional units of MTP2. In the first place, it coordinates the transfer of signal units. LC also monitors the operation of the signaling link. It communicates

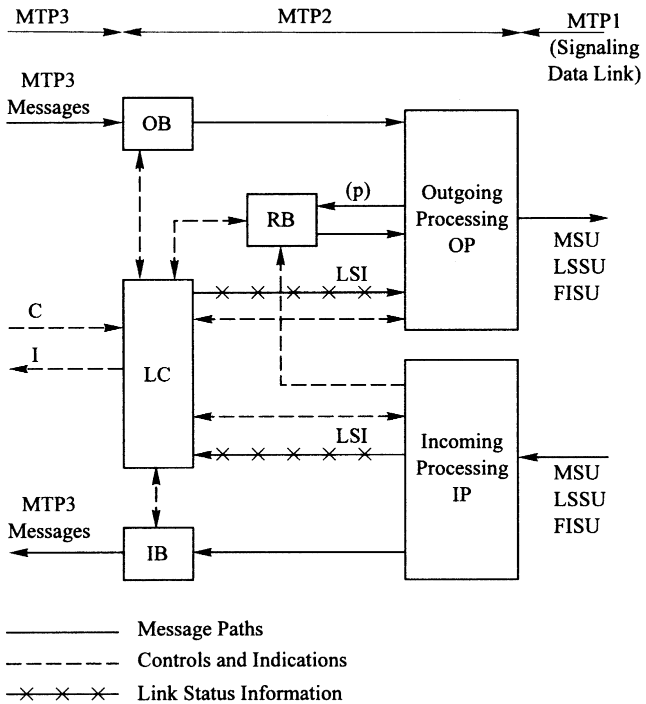


Figure 8.3-1. Structure of MTP2.

with its MTP3, accepting link status commands (C) and reporting link status information with indications (I). Finally, LC communicates with the LC at the distant end of the signaling link, using link status signal units.

The MTP3 in a signaling point places its outgoing MTP3 messages in the *output buffer* (OB) of the signaling link. *Retransmission buffer* (RB) stores messages that have been sent but have not yet been positively acknowledged by the distant MTP2.

Each message to be transmitted or retransmitted passes through *outgoing processing* (OP) and then enters the signaling data link as a MSU. A signal unit received from the signaling data link is processed by *incoming processing* (IP). The MTP3 messages in the MSUs that are accepted by IP are placed in *input buffer* (IB) and are retrieved by MTP3.

All buffer transfers are "*first in, first out*": a MTP2 takes outgoing messages from its output buffer in the same order in which they were placed there by MTP3, and MTP3 takes received messages from the input buffer of a MTP2 in the same order in which they were entered by MTP2. This is one of the requirements for in-sequence MSU delivery.

8.3.2 Outgoing Processing

In this section, we examine the MTP2 actions in transmitting a MSU, with the aid of Figs. 8.3-1 and 8.3-2.

A working signaling link transfers uninterrupted streams of signal units in both directions. When the transmission of a SU has been completed, LC selects the next SU to be sent out. LSSUs have the highest priority. When no LSSUs have to be sent, the oldest message in OB or RB is selected for transmission.

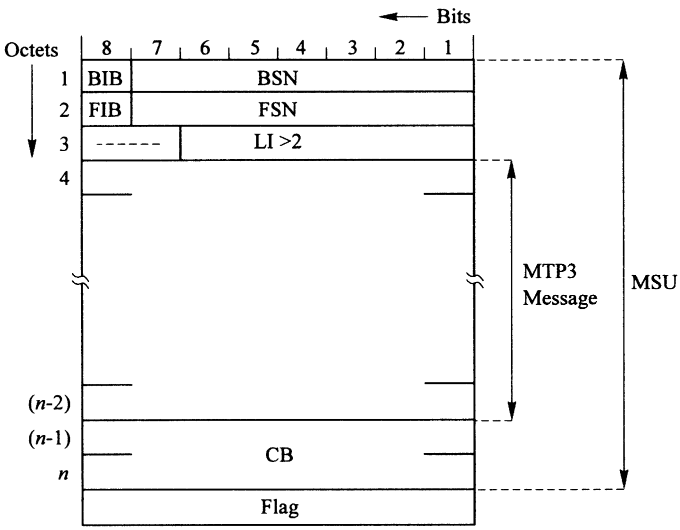


Figure 8.3-2. Parameters added and processed by MTP2. (From Rec. Q.703. Courtesy of ITU-T.)

The rules for selecting a message from OB or RB depend on the type of error control and are discussed in Sections 8.4 and 8.5. When no LSSU or message is awaiting transmission, fill-in signal units (FISUs) are sent.

MTP2 Parameters. Outgoing processing adds a number of MTP2 parameters to the message (Fig. 8.3-2).

Forward Sequence Number (FSN). A forward sequence number is assigned to a message when it is removed from the OB for its initial transmission. The FSN field has seven bits, and the sequence numbers are therefore confined to the range 0–127. The numbers are assigned cyclically to consecutive messages: 0, 1, 2, . . . , 126, 0, 1, . . . , and so on.

Backward Sequence Number (BSN). This number indicates the FSN of the most recently accepted incoming MSU.

Forward and Backward Indicator Bits (FIB, BIB). These are discussed in Section 8.4.

Length Indicator (LI). This indicates the length (number of octets) of the message.

Check Bit (CB) Field. As in SS6 signaling, error detection uses cyclic redundancy checking (see Section 5.2.3). The SS7 checking algorithm is a modulo 2 division of the contents in octets 1 through $(n-2)$ of the MSU by the 17-bit divisor 1 0001 0000 0010 0001. The remainder of the division is placed in the CB field.

Flag (F). Consecutive signal units on a signaling data link are separated by a flag octet F, coded 01111110.

We now consider the outgoing processing of a MTP3 message from OB (Fig. 8.3-1):

1. The message is moved from OB to OP, where its length (LI) is determined and a FSN is assigned.
2. A copy of the message, including its LI and FSN, is sent to RB (path P), where it remains until a positive acknowledgment is received.
3. The values of BSN, FIB, and BIB are determined.
4. The contents of CB are determined.
5. Zero insertion. All SUs are separated by flags. Since the 01111110 flag pattern may also occur in the MSU (octets 1 through n), the MSU is scanned, and a "zero" is inserted behind each string of five consecutive "ones." In this way, the 01111110 pattern never occurs inside the MSU. The MSUs on a signaling data link thus are not always completely identical to the MSU shown in Fig. 8.3-2, because of possible inserted zeros.
6. A flag is appended to the MSU, which then enters the signaling data link.

8.3.3 Incoming Processing of Signal Units

Incoming processing of received SUs (MSUs, LSSUs, and FISUs) is described below, again with the aid of Figs. 8.3-1 and 8.3-2.

1. Zero deletion. Having recognized the closing flag of the previous signal unit, IP scans the subsequent incoming bits, until it encounters the next flag. During the scan, it removes all zeros that follow five consecutive ones. After this, the SU consists again of an integral number of octets.
2. Error detection. This consists of modulo 2 division of the contents of octets 1 through (n – 2) by the divider polynomial. The result is compared against the received CBs. All SUs with errors are discarded. Error-free SUs are separated by type (LSSU, FISU, MSU—derived from the value of LI), and acceptance criteria depend on the SU type.
3. Error-free MSUs are "sequence screened." The basic criterion for acceptance of a MSU is that its FSN should exceed the FSN of the most recently accepted MSU by 1 (modulo 128), because this indicates that the MSU has been received in-sequence. This step is part of the procedure that maintains in-sequence delivery of MSUs when transmission errors occur.
4. IP processes the parameters BSN and BIB of accepted MSUs, which play a role in error correction (Sections 8.4 and 8.5).

8.3.4 Other Outgoing and Incoming Processing

So far, we have discussed the processing steps for the initial transmission and the acceptance of MSUs. We now examine the processing in other cases.

Retransmitted MSU. When a message that has already been transmitted needs to be retransmitted, the message and parameters LI and FSN are copied from RB. The retransmitted MSU thus has the same FSN value as the originally transmitted one. This is part of the procedure for in-sequence MSU delivery when transmission errors occur. Outgoing processing determines the values of FIB, BSN, and BIB and then does steps 3 through 6 of Section 8.3.2.

Retransmitted MSUs remain in RB until they have been positively acknowledged.

Transmission of LSSUs and FISUs. Outgoing LSSUs and FISUs originate at link control LC. They are processed by outgoing processing and therefore include the parameters FSN, FIB, BSN, and BIB (Fig. 8.3-3). However, OP does not assign a "new" FSN value to these signal units; they receive the FSN value of the most recently transmitted MSU. Outgoing LSSUs and FISUs are never retransmitted, and link control periodically retransmits a LSSU until it receives a response from the LC in the distant MTP2. The parameters FSN, FIB, BSN, and BIB are processed by incoming processing only if the LSSU or FISU is error-free and in-sequence (its FSN matches the FSN of the most recently accepted MSU). Link status information (parameter LSI) in error-free LSSUs is always passed to LC for processing.

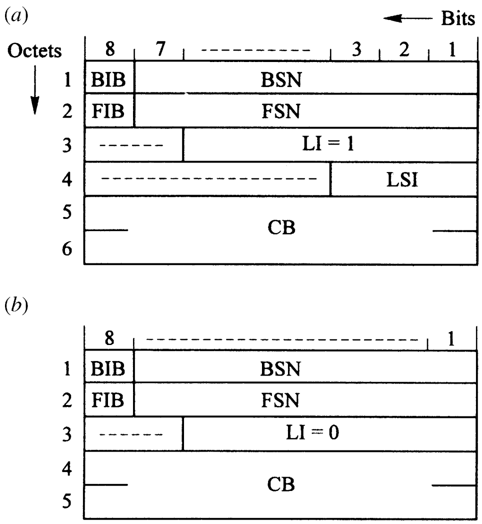


Figure 8.3-3. (*a*) Link status unit and (*b*) fill-in signal unit. (From Rec. Q. 703. Courtesy of ITU-T.)

8.4 BASIC ERROR CORRECTION

8.4.1 Introduction to Error Correction

Error correction of MSUs consists of screening received error-free MSUs for acceptance, positively acknowledging accepted MSUs and retransmitting MSUs that have not been accepted by the distant MTP2.

ITU-T has defined two error-correction procedures for SS7 [7]. Basic error correction, described in this section, is used on signaling links whose lengths do not exceed some 8000 km. Preventive cyclic retransmission, which is used on longer links, is discussed in Section 8.5.

We examine error correction of MSUs sent by MTP2-A on the signaling link of Fig. 8.1-1. The same procedure is used for MSUs sent in the opposite direction.

8.4.2 Actions at the MTP2s

Acknowledgments. Basic error correction includes *positive* and *negative* acknowledgments of received MSUs. The acknowledgment information for MTP2-A is in the BSN (backward sequence number) and BIBs (backward indicator bits) of SUs (of any type) sent by MTP2-B. In both positive and negative acknowledgments, the BSN is equal to the FSN (forward sequence number) of the MSU that has been most recently accepted by MTP2-B. Positive and negative acknowledgments are indicated by the BIBs in consecutive SUs. In Fig. 8.4-1, SU_2 has the same BIB value as SU_1. This signifies the positive acknowledgment of MSUs with FSN up through 26. SU_3 indicates a negative acknowledgment, because its BIB differs from the BIB of SU_2.

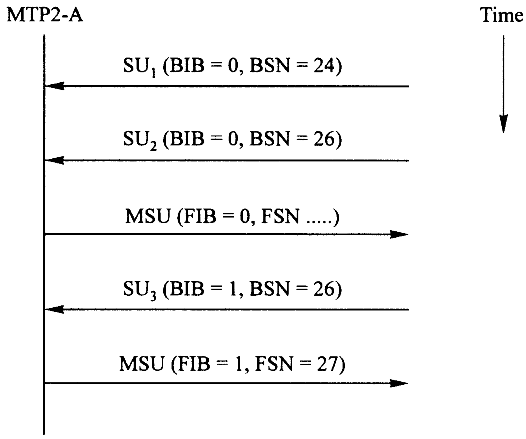


Figure 8.4-1. Positive and negative acknowledgments.

Response to Acknowledgments. MTP2-A can be in one of two transmission states: transmitting MSUs from its output buffer, or retransmitting previously sent MSUs from its retransmission buffer. When receiving a positive acknowledgment, MTP2-A removes the acknowledged MSUs from its RB and remains it its current state. On receipt of a negative acknowledgment, it starts (or restarts) a retransmission cycle, beginning with the MSU in its RB whose FSN exceeds the FSN of the most recent positively acknowledged MSU by 1 (modulo 128).

In Fig. 8.4-1, after receipt of the negative acknowledgment in SU_3, MTP2-A starts retransmitting MSUs, beginning with the MSU that has FSN = 27.

The FIBs of the SUs sent by MTP2-A are copied from the BIB of the most recently received SU. For example, in the MSU sent after receipt of SU_3, FIB is set to 1.

Acceptance of Received MSUs. All SUs received at MTP2-B input processing are checked for errors, and SUs with errors are discarded. Input processing subjects error-free MSUs to two additional screening steps. It first compares the FIB of the received MSU with the BIB it sent in its latest SU. If the FIB is not equal to the BIB, then MTP2-A has not yet received the latest negative acknowledgment from MTP2-B, and the MSU is discarded. If FIB = BIB, the MSU is sequence-checked. It is accepted if its FSN exceeds the FSN of the most recently accepted MSU by 1 (mod 128). If the check passes, MTP2-B accepts the MSU (places it in its input buffer) and positively acknowledges the MSU in the next sent SU. If the sequence check fails, it discards the MSU and includes a negative acknowledgment in its next SU.

8.4.3 Error Correction Example

An example of basic error correction for MSUs sent by MTP2-A is shown in Fig. 8.4-2. The LSSUs and FISUs from MTP2-A are not shown, because they play no role in the error correction of the sent MSUs. The time required by the SUs to traverse the signaling link is indicated by the sloping lines.

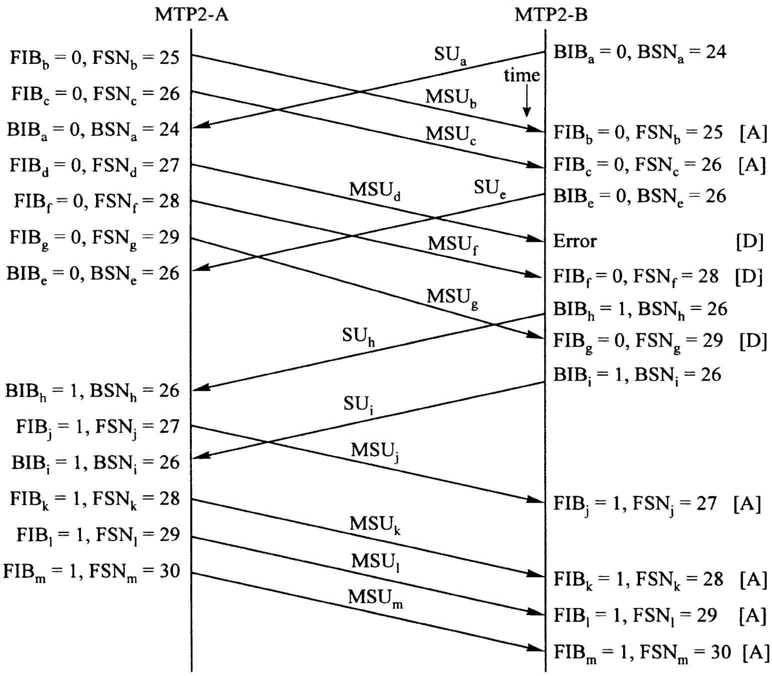


Figure 8.4-2. Example of basic error correction of MSUs sent by MTP2-A. [A], MSU is accepted; [D], MSU is discarded.

The values of FSN and FIB in the MSUs from MTP2-A, and of BSN and BIB in SUs (MSUs, LSSUs, FISUs) from MTP2-B, are shown at both MTPs. This makes it easier to correlate the figure and the text below.

We assume that, prior to the transmission of SU_a, MTP2-B has been accepting MSUs and has been sending SUs with BIB = 0. MTP2-A has been sending MSUs from its output buffer, with FIB = 0.

SU_a positively acknowledges MSUs with FSN up through 24. MTP2-A keeps sending MSUs. MSU_b and MSU_c pass the tests at MTP2-B and are accepted, because FIB_b and FIB_c are equal to BIB_a, and the FSNs indicate the proper sequence. SU_e positively acknowledges the MSUs in SU_e ($BIB_e = BIB_a$ and $BSN_e = 26$).

MTP2-A keeps transmitting MSUs from its output buffer. MSU_d incurs a transmission error and is discarded by MTP2-B incoming processing. MSU_f arrives without error, and with the proper FIB value, but fails the sequence test. It is therefore discarded and negatively acknowledged by SU_h ($BIB_h \neq BIB_e$, $BSN_e = 26$).

MSU_g is discarded because its FIB_g does not match BIB_h. In this situation, MTP2-B does not send another negative acknowledgment, but signals a positive acknowledgment in SU_i ($BIB_i = BIB_h$), in which BSN_i indicates that MSU_c (with FSN = 26) is still the latest accepted MSU.

MTP2-A receives the negative acknowledgment in SU_h. Since $BSN_h = 26$, it starts a retransmission cycle of MSUs in its RB, beginning with MSU_j

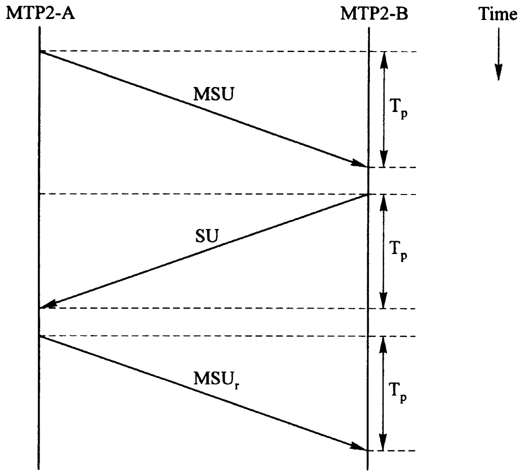


Figure 8.4-3. Transfer delay of a retransmitted MSU.

(retransmission of MSU_d, which has $FSN_d = 27$), and setting FIB to 1. The receipt of the positive acknowledgment in SU_i does not change things at MTP2-A, which continues its retransmission cycle until an MSUs in its retransmission buffer have been sent again.

MTP2-B accepts MSU_j, MSU_k, and MSU_l, because their FIBs are equal to BIB_i, and their FSNs indicate in-sequence delivery.

When MSU_l (retransmission of MSU_g) has been sent, the retransmission cycle is complete, and MTP2-A resumes the initial transmission of MSUs in its output buffer.

8.4.4 Message Transfer Delays

We now consider the initial transmission of a MSU, sent by MTP2-A. It arrives at MTP2-B after a propagation delay (T_p) (Fig. 8.4-3). Suppose that the MSU is rejected by MTP2-B and negatively acknowledged in a signal unit (SU). This signal unit arrives at MTP2-A after a delay of at least $T_{p.}$

MSU_r, the retransmitted copy of MSU, arrives at MTP2-B with a delay of at least $3T_p$ (measured from the time when the original MSU left MTP2-A).

Since SS7 call-control applications are real-time critical, basic error correction is used only on signaling links, which have propagation times (T_p) that do not exceed 40 ms [1]. This corresponds to signaling links with lengths below about 8000 km.

8.5 PREVENTIVE CYCLIC RETRANSMISSION

8.5.1 Introduction

Preventive cyclic retransmission (PCR) is designed for use on signaling links with large propagation times (T_p), for example, signaling links that are carried on satellite circuits (Section 1.5.4). As in basic error correction, the FSN identifies the position of a MSU in its original sequence of transmission, and the BSN always identifies the most recently accepted message signal unit.

PCR uses positive acknowledgments only. Indicator bits FIB and BIB are ignored (they are permanently set to "1"), and incoming processing simply accepts or discards an error-free MSU based on the value of its FSN, which has to exceed the FSN of the most recently accepted MSU by one unit (modulo 128).

8.5.2 Preventive Retransmission Cycles

Whenever there is no LSSU or new MSU to be sent out, MTP2-A, in lieu of sending a FISU, starts a *preventive retransmission* cycle in which the MSUs in its retransmission buffer are retransmitted in-sequence and starting with the oldest One (lowest FSN). The retransmitted copies of MSUs that have already been accepted by MTP2-B now arrive out-of-sequence and are discarded. However, if any of the original MSUs were not accepted, their retransmission takes place much sooner than in basic error correction (where a time of at least $2T_p$ is required for the negative acknowledgment to reach MTP2-A). This is why PCR can be used on signaling data links with propagation times that make basic error correction impractical.

A preventive retransmission cycle ends in one of two ways. If all MSUs in RB have been retransmitted and no LSSU or new MSU has to be sent out, MTP2-A starts sending FISUs. A retransmission cycle also ends when a LSSU or new MSU has to be sent.

8.5.3 Forced Retransmission Cycles

Under normal traffic loads, about 20% of the octets on a signaling link with PCR carry MSUs that are transmitted for the first time. This provides ample opportunity to start and complete preventive retransmission cycles. However, during bursts of high MSU volume, preventive cyclic retransmissions may not occur as often as needed. Therefore, PCR also includes *forced* retransmission cycles.

Link control (LC) in a MTP2 (Fig. 8.3-1) constantly monitors the number of message signal units (N_1) and the number of octets (N_2) in its RB. If either of these numbers exceeds a predetermined threshold value, a *forced retransmission cycle* is initiated. This type of retransmission cycle ends only after all MSUs in the RB have been retransmitted.

8.5.4 Comparison with Basic Error Correction

As has been mentioned, PCR is used on signaling links with propagation times in excess of some 40 ms, because basic error correction on such links results in MSU queuing delays that are unacceptable for call-control applications (TUP, ISUP).

On the other hand, basic error correction is preferred on signaling links with propagation times below 40 ms, because it allows higher MSU loads on the signaling links than PCR [7].

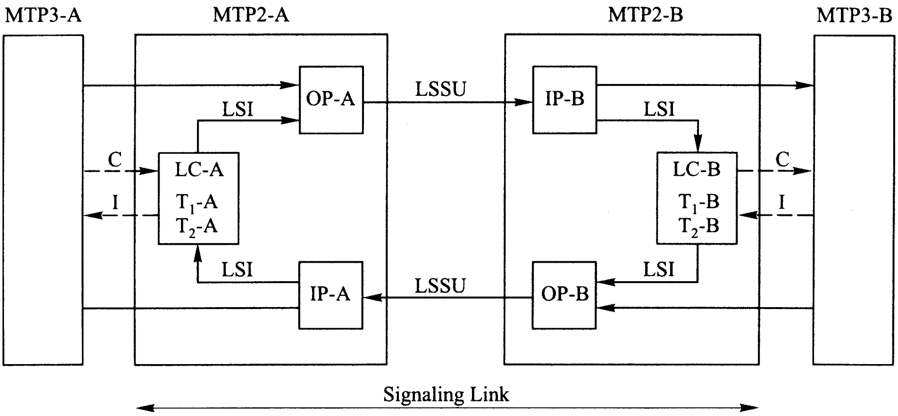


Figure 8.6-1. Transfer of link status information (LSI). LSSU, Link status signal unit; IP, incoming processing; OP, outgoing processing; LC, link control; T_1, T_2, timers; C, commands from MTP3; I, indications to MTP3.

8.6 SIGNALING LINK MANAGEMENT

Signaling link management responsibilities at a signaling point are shared by the MTP2s of the individual signaling links and MTP3. This section describes the MTP2 signaling link management functions [7]. The MTP2 management functions monitor the status of the signaling link and, when necessary, pass status indications to MTP3.

We consider a signaling link between signaling points A and B (Fig. 8.6-1). The management functions are performed by link controls (LC-A and LC-B). Each LC communicates with the MTP3 at its signaling point, accepting controls (C) and sending status indications (I).

LC-A and LC-B also send link status information (LSI) to each other. LSI originated by LC-A is embedded in a link status signal unit (LSSU) by OP-A (outgoing processing at A). IP-B (incoming processing at B) extracts the LSI and passes it to LC-B. Transfer of LSI in the other direction is done in the same manner.

The LSI is a parameter in a LSSU—see Fig. 8.3-3(a). The other LSSU parameters are the same as the MTP parameters in message signal units. Length indicator (LI) = 1 indicates that the signal unit is a LSSU. The link status information is coded as follows:

LSI Bit: 321	Acronym	Link Status
000	SIO	Out of alignment
001	SIN	Normal alignment
010	SIE	Emergency alignment
011	SIOS	Out of service
100	SIPO	Processor outage
101	SIB	Busy

In the link management examples that follow, a LSSU with a particular value of LSI is denoted by the corresponding acronym. For example, a "SIN" denotes a LSSU in which LSI = 001.

8.6.1 Initial Alignment

When a signaling link is turned on, both LCs start sending SIOs. When LC-A acquires alignment (i.e., recognizes the flags in the incoming bit stream), it starts sending SINs. When LC-A begins to receive SINs, indicating that LC-B also has achieved alignment, a proving period of a few seconds is started. If both LCs maintain their alignment during this period, they start sending FISUs and indicate to their respective MTP3s that the link is *aligned and ready* for service.

If LC-A has acquired alignment but keeps receiving SIOs or SIOSs, it knows that LC-B is not in working condition. LC-A then passes an *alignment failure* indication to MTP3-A.

8.6.2 Error Monitoring

When the signaling link is in service, each LC monitors the error rate of received signal units. When one of the following conditions occurs, the MTP3 in the signaling point is alerted with a *link failure* indication:

1. Sixty-four consecutive signal units have been received with errors.
2. The error rate of received signal units exceeds one error per 256 signal units.
3. An "impossible" bit pattern, consisting of at least seven consecutive 1's, has been received, and a flag has not been detected within 16 octets following that pattern.

8.6.3 Delayed Acknowledgments

Each LC expects that, under normal conditions, its outgoing MSUs are acknowledged within a short time. Excessive delay of acknowledgments is an indication of problems on the signaling link. Acknowledgment delays are checked by both LCs, using timers (T_1), which expire in 0.5–2 s (Fig. 8.6-1).

The procedure at MTP2-A is as follows. Timer T_1-A is restarted each time an acknowledgment with a new value of BSN (Section 8.3) is received from MTP2-B, and there is at least one MSU (waiting for acknowledgment) in the RB. If T_1-A expires, LC-A interprets the absence of received acknowledgments during the time-out interval as a signaling link problem and passes a *link failure* indication to its MTP3-A.

8.6.4 Level 2 Flow Control

When a link control, say, LC-A, detects that the number of received MSUs in its IB exceeds a particular value, because MTP3-A has fallen behind in taking out these

MSUs, it starts sending SIB (link busy) status units to LC-B, at intervals of 80–120 ms. It continues sending outgoing MSUs and FISUs but discards incoming MSUs and "freezes" the value of BSN in the SUs it sends out. The delay in acknowledgments would normally cause timer (T_1-B) of LC-B to time out. However, timers (T_1) are also restarted each time a SIB is received from the distant MTP2. Therefore, T_1-B does not expire as long as SIBs are being received.

When LC-B receives the first SIB, it also starts a timer T_2-B, which expires in 3–6 s. If the congestion at MTP2-A abates, it again acknowledges received signal units, and MTP2-B starts to receive SUs with "new" values of BSN. If this happens before T_2-B has expired, LC-B stops the timer and resumes normal operation. However, if T_2-B expires and LC-A is still sending SUs with the "frozen" BSN value, LC-B passes a *link failure* indication to its MTP3-B.

8.6.5 Processor Outage

With this procedure, the MTP3 at either end of the signaling link can temporarily suspend the operation of the link. Assume that LC-A in Fig. 8.6-1 has received a command (C) from MTP3-A to suspend incoming and outgoing message traffic on the link. LC-A then starts sending a continuous stream of SIPOs (processor outage) and discards received MSUs. On receipt of the SIPOs, link control (LC-B) starts sending fill-in signal units (FISUs)—see Fig. 8.3-3(*b*)—and indicates to its MTP3-B that the signaling link is out of operation.

When LC-A receives a command from its MTP3-A to resume operation, it stops sending SIPOs and resumes sending MSUs and FISUs. LC-B, which no longer receives SIPOs, indicates to its MTP3-B that the signaling link is back in operation and resumes sending MSUs.

8.6.6 Outgoing Congestion

Link controls (LCs) constantly monitor N, the total number of outgoing MSUs in their output and retransmit buffers, OB and RB (Fig. 8.3-1). When N exceeds a threshold value N_{on}, LC passes an *onset of congestion* indication to its MTP3, which then takes measures to reduce the outgoing MSU traffic temporarily. When N drops below a threshold value N_{off}, the link control passes an *end of congestion* indication, and MTP3 resumes normal outgoing message flow. To avoid rapid fluctuations in link status, N_{off} is set below N_{on}. This gives a certain amount of hysteresis in the status transitions.

8.7 OVERVIEW OF MTP LEVEL 3

The message transfer part level 3 (MTP3) is divided into the fol1owing two groups of functions (Fig. 8.7-1).

Signaling Message Handling (SMH). This handles the transfer of messages between "peer" MTP-users–Telephone User Part (TUP), Integrated Services

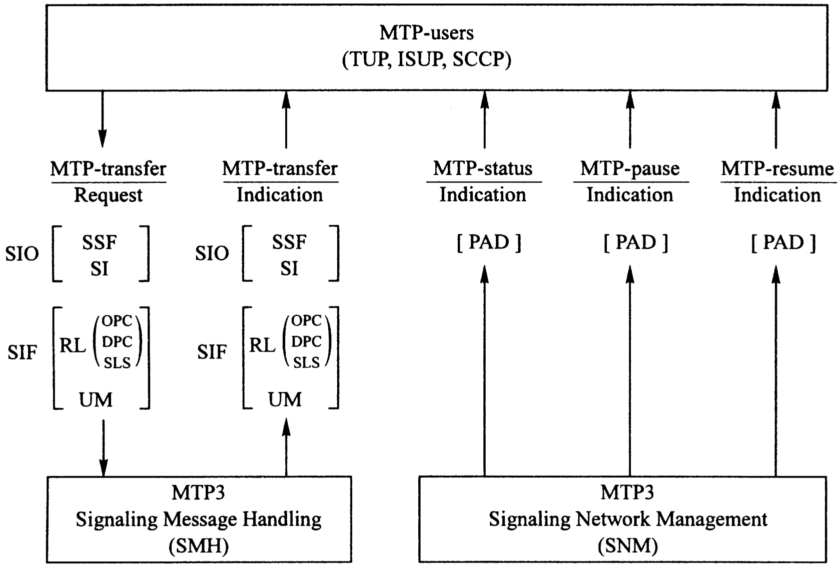


Figure 8.7-1. MTP3 primitives. SIO, Service information octet; SSF, sub service field; SI, service indicator; SIF, signaling information field; RL, routing label; OPC, origanting point code; DPC, destination point code; SLS, signaling link section; UM, user message; PAD, point code of affected destination. (From Rec. Q.701 and Q.704. Courtesy of ITU-T.)

Digital Network User Part (ISUP), and Signaling Connection Control Part (SCCP). SMH is described in Section 8.8.

Signaling Network Management (SNM). This keeps the message traffic flowing under abnormal conditions (congestion, failures) in the signaling network. SNM is described in Section 8.9.

8.7.1 MTP-Transfer Primitives

MTP-users pass their outgoing messages to SMH in MTP-transfer requests, and SMH delivers incoming messages to the MTP-users in MTP-transfer indications (Section 7.3.2).

The primitives include the parameters in the service information octet (SIO) and the signaling information field (SIF) (Fig. 7.3-3). The parameters are discussed in Section 8.8.

8.7.2 MTP Status, Pause, and Resume Indications

SNM passes these indications to the MTP-users at its signaling point, to inform them about abnormal conditions (failures, congestion) in the signaling routes to particular destinations. The affected destinations are identified by their point codes.

8.8 MTP3 SIGNALING MESSAGE HANDLING

8.8.1 Message Format

The general format of MTP3 messages is shown in Fig. 8.8-1. We distinguish *service information octet* (SIO) and the *signaling information field* (SIF). In turn, SIF is divided into the *routing label* (RL) and the *user message* (UM). All information originates at the MTP-user, that sends the message and is included in a MTP-transfer primitive (Fig. 8.7-1). MTP3 examines the data, in SIO and RL. However, the user message is transferred transparently (without examination).

8.8.2 Parameters in Routing Label

The parameters in the RL are used by the MTP3s in the signaling points along the message path to determine the signaling route set to the message destination. The RL

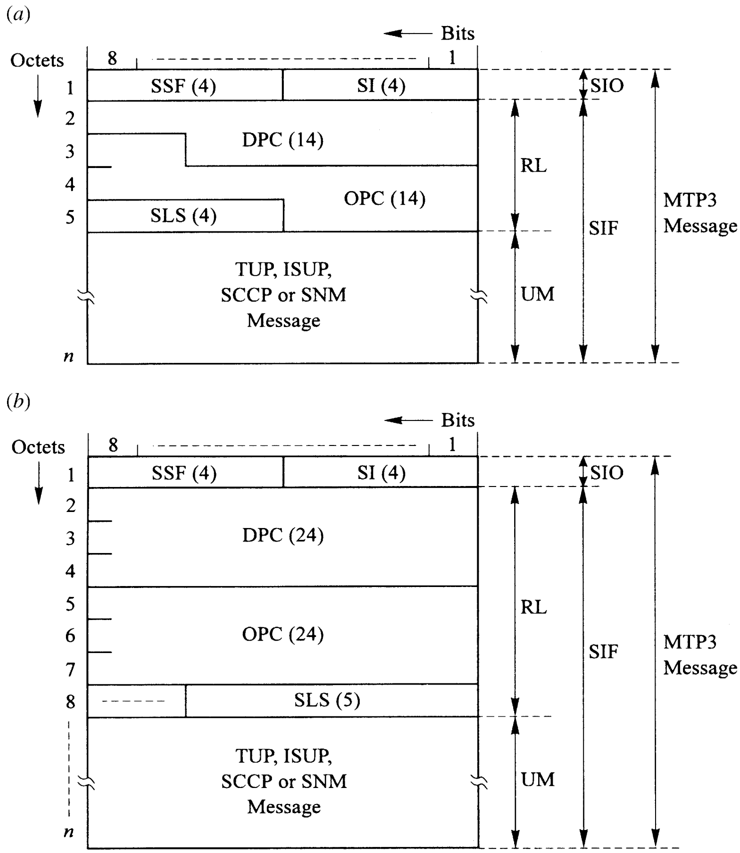


Figure 8.8-1. MTP3 message format: (*a*) ITU-T SS7 and (*b*) ANSI No. 7. (From Rec. Q.704. Courtesy of ITU-T.)

formats of ITU-T SS7 and ANSI No. 7 are slightly different—see Fig. 8.8-1—because of differences in parameter lengths [5,9].

Destination Point Code (DPC) and Originating Point Code (OPC). These parameters identify the point code of the destination and originating signaling points, respectively.

The point codes in ITU-T SS7 and ANSI No. 7 have 14 and 24 bits, respectively (Fig. 7.2-1).

Signaling Link Selector (SLS). This parameter divides the outgoing message loads of the MTP-users in a signaling point into 16 or 32 approximately equal parts. SLS is used for the selection of a particular signaling link in the signaling route set to the message destination (Section 8.8.5).

8.8.3 Parameters in Service Information Octet

These parameters are used by the MTP3 in the destination: signaling point to deliver the message to the appropriate MTP-user.

Service Indicator (SI). The code in of SI represents the MTP-user. For the international network, ITU-T has specified the following SI codes:

SI	MTP-User
0000	SNM
0001	Signaling network testing and maintenance
0011	SCCP
0100	TUP
0101	ISUP

The MTP-users represented by SI = 0000 and 0001 have not been mentioned before and are described in Section 8.9.

Subservice Field (SSF). Two SSF codes have been defined by ITU-T:

SSF	
0000	International network
0010	National network

The SSF field is of importance in *international switching centers* (ISCs), which belong to both the international network and a national network. Since a MTP-user protocol, say, TUP, for the international network is usually somewhat different from the same protocol in a national network, an ISC is equipped with two TUP versions, and SSF specifies the version to which the message is to be delivered.

8.8.4 SMH Functions

Signaling message handling functions are divided into three groups (Fig. 8.8-2).

SMH Message Discrimination. This function group uses the value of SSF and DPC to determine whether a message received from a MTP2 is destined for its signaling point. The value of SSF indicates whether DPC contains a national or international point code.

When SMH message discrimination recognizes the DPC as the point code of its signaling point, it passes the message to SMH message distribution. When the message has another destination, the message handling depends on the type of signaling point. If the signaling point is a signal transfer point (STP) or a combined transfer and end point (STEP) (Section 5.1), SMH message discrimination passes the message to SMH message routing. If the signaling point is a signaling end point (SEP), the message is discarded.

SMH Message Distribution. This function group delivers the messages received from SMH message discrimination to the MTP-user specified by SI and SSF.

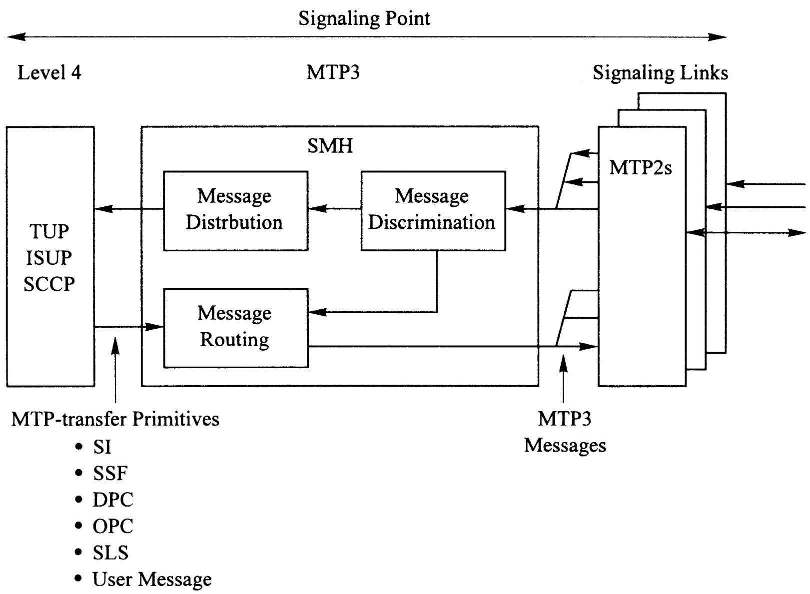


Figure 8.8-2. Structure and interfaces of signaling message handling (SMH). (From Rec. Q.704. Courtesy of ITU-T.)

SMH Message Routing. This function group selects the outgoing signaling links for outgoing messages (received from the MTP-users in the signaling point) and transit messages (received from SMH message discrimination).

8.8.5 Selection of Outgoing Signaling Links

SMH message routing first determines the signaling route set for the message, based on the values of SSF and DPC, which indicate the point code of a national or international destination.

The selection of a particular outgoing signaling link in a route set has two objectives. In the first place, a signaling point should distribute its messages to a particular destination equally among the signaling links in its route set to that destination.

In the second place, all messages pertaining to a particular trunk, or a particular transaction, should be routed on the same signaling route. This is necessary to avoid out-of-sequence delivery of messages pertaining to a trunk, or a transaction. We have seen that SS7 messages are delivered in-sequence from the MTP2 at one end of the link to the MTP2 at the other end (Section 8.3). However, messages spend some "queueing delay" time in the output and input buffers of the MTP2s at the ends of a signaling link. These delays vary from moment to moment and from link to link. If two messages, M_1 and M_2, which pertain to a trunk or transaction would traverse different signaling routes, inequalities in the delays on the signaling links along routes could result in out-of-sequence message delivery.

Therefore, the MTP-users assign a particular SLS value to all outgoing messages relating to a particular trunk or transaction, and SMH message routing determines the outgoing signaling link in a route set from the value of SLS.

The assignments of SLS can be made in several ways. For example, in TUP and ISUP, each trunk in a trunk group is identified by a circuit indication code (CIC)—see Section 7.2.5—and SLS values can be derived from CIC codes as follows:

$$\text{ITU-T SS7:} \quad \text{SLS} = \text{CIC mod } 16$$
$$\text{ANSI No. 7:} \quad \text{SLS} = \text{CIC mod } 32$$

Since each SLS value covers approximately the same number of trunks, the loads of outgoing messages the various SLS values are approximately equal.

We shall see in Chapter 15 that SCCP can provide in-sequence delivery of messages relating to individual transactions, by assigning the same SLS value to all outgoing messages of a particular transaction.

A number of link-selection examples are outlined below.

Example 1. Figure 8.8-3(*a*) shows the signaling links in the route set between signaling points A and F in the original Bell System signaling network (see Section 5.1). We examine the signaling link selection for messages from A to F [10]. Under normal (nonfailure) conditions in the signaling route set, signaling links SL_9 and SL_{10} are not used.

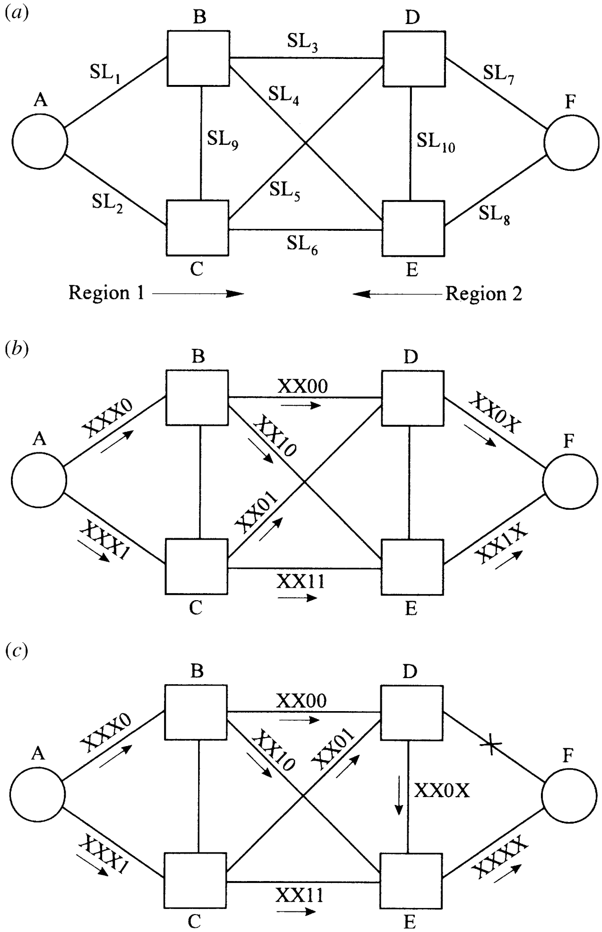


Figure 8.8-3. Load sharing examples: (*a*) signaling route set for messages from A to F, (*b*) load sharing under normal conditions, and (*c*) load sharing on failure of SL_7. (From Rec. Q.705. Courtesy of ITU-T.)

We assume a four-bit SLS field and denote the bits in this field by

$$\text{SLS:} \quad s_4 \ s_3 \ s_2 \ s_1$$

The route sets to destination F at signaling points A, B, and C contain two signaling links each. We assume that SMH message routing at originating signaling points uses bit s_1 of SLS to select the outgoing signaling link. In this example, messages from A with $s_1 = 0$ and $s_1 = 1$ are sent on, respectively, SL_1 and SL_2.

Signal transfer points base their selections on bit s_2. In this example, we assume that signal transfer points B and C select, respectively, SL_3 and SL_5 when $s_2 = 0$, and SL_4 and SL_6 when $s_2 = 1$. Formally, signal transfer points D and E also use

s_2 to make the outgoing link selection. However, D and E have only one normal route to F, which is used for both values of s_2.

Figure 8.8-3(*b*) displays the SLS codes in the messages carried by the individual signaling links (bits that can be 0 or 1 are marked X). The links SL_1, SL_2, SL_7, and SL_8 carry messages with eight of the 16 SLS codes each (about 50% of the messages). Also, each of the signaling links SL_3, SL_4, SL_5, and SL_6 carries messages with 4 SLS codes (about 25% of the messages).

Under this form of signaling link selection, the signaling route for messages with a particular value of SLS, and sent by signaling point A, can be different from the route for messages with the same SLS that are sent by signaling point F.

Example 2. Suppose that SL_7 in Fig. 8.8-3(*a*) has failed. Signal transfer point D thus has to divert the messages with destination F to SL_{10}. Link SL_{10} then carries the messages with SLS = XX0X, and SL_8 carries the entire message load—see Fig. 8.8-3(*c*).

Example 3. In some signaling networks, normal signaling routes pass through more than two signal transfer points (STPs). Figure 8.8-4 shows an example that occurs in the postdivestiture North American signaling network [1]. Exchange A and signal transfer points B and C are owned by a local exchange carrier, while exchange H and signal transfer points D, E, F, and G belong to the signaling network of an interexchange carrier (Section 1.1.3). We examine the distribution of messages from exchange A to exchange H over the various signaling links.

The outgoing link selections at A, B, C, D, E, F, and G are based on the values of bits in the five-bit (ANSI) signaling link selector SLS field. For equal message load division, the selections of outgoing signaling links at the signaling points (SPs) along the route have to be based on different bits of SLS. In this example, signaling point A, which is the originating SP in the signaling route, uses bit s_{17}. STPs B and C are the second SPs in the routes from A to H and therefore use bit s_2. STPs D and E are the third SPs in the routes and thus use bit s_3. STPs F and G are the fourth signaling points in the routes and formally base their selection on the value of s_4. However, F and G have only one normal signaling route to destination H, which is used for both values of s_4. Figure 8.8-4 shows the SLS codes on the individual signaling links, for messages sent by A.

Since the STPs cannot determine whether they are the second, third, or fourth SP in the signaling route of an incoming MSU, the link selection procedure is implemented as follows. All signaling points use the rightmost bit of SLS for outgoing link selection and then *rotate* the contents of SLS one position to the right.

In this way, signaling point A uses bits s_1, STPs B and C use s_2, STPs D and E use s_3, and F and G (formally) use s_4 of the *original* SLS code (assigned by the originating MTP-user at signaling point A). For simplicity, the SLS codes on the signaling links in Fig. 8.8-4 are shown "without rotation" (s_1 always appears in the rightmost position of SLS). The actual codes on the links can be derived easily. For example, the SLS code in the messages on the signaling link between B and D is shown as XXX00. The actual code has undergone two rotations and is therefore 00XXX.

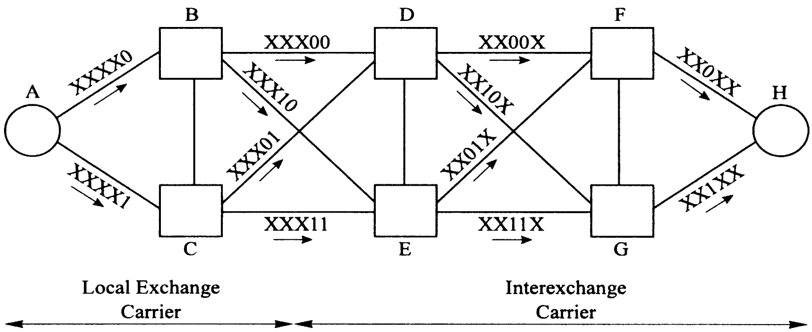


Figure 8.8-4. Load sharing for internetwork messages from A to H. (From *IEEE Commun. Mag.* **28**(7). Copyright © 1990 IEEE.)

8.9 MTP3 SIGNALING NETWORK MANAGEMENT

The purpose of MTP3 signaling network management is to keep the signaling message traffic flowing under abnormal (congestion, failure) conditions in the signaling network. Some of these conditions may require a temporary reduction—or suspension—of outgoing message traffic to certain destinations. In these cases, SNM alerts the level-4 protocols at its signaling point [8,10].

8.9.1 Structure and Interfaces of SNM

SNM consists of three parts—see Fig. 8.9-1.

SNM Link Management. This monitors and controls the status of the individual MTP2s (signaling links) of the signaling point.

SNM Route Management. This communicates with its "peer" functions at other signaling points, sending and receiving information regarding the status of signaling routes to individual destinations.

SNM Traffic Management. This receives information from SNM link management on the status (available/unavailable) of the signaling links at its signaling point, and information from SNM route management about problems on the signaling route sets to particular destinations. When necessary, it informs SMH message routing and the MTP-users at its signaling point.

Some SNM procedures involve more than one part of SNM and require communications between these parts.

Interfaces of SNM. SNM has interfaces with the MTP-users (level-4 protocols), the MTP2s of the signaling links, and the SMH message routing function at its

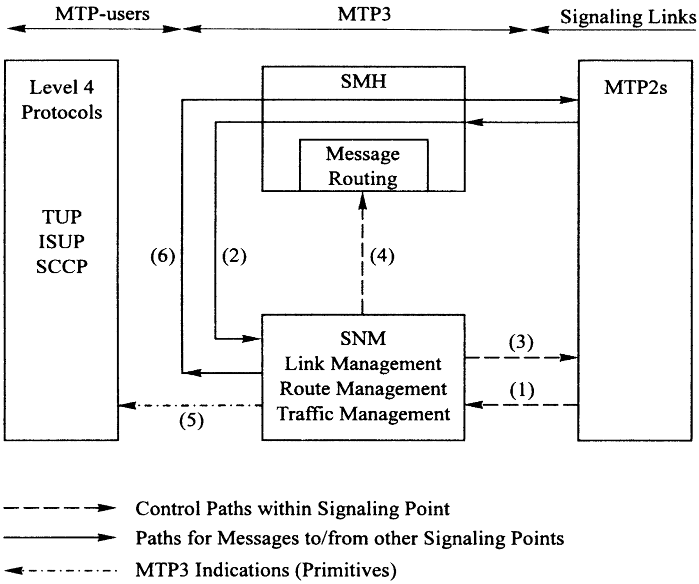


Figure 8.9-1. Structure and interfaces of signaling network management (SNM). (From Rec. Q.704. Courtesy of ITU-T.)

signaling point. It also communicates with SNMs at other signaling points, sending and receiving *SNM messages*.

Inputs. SNM bases its actions on the following inputs (the numbers below correspond to the numbers in Fig. 8.9-1):

1. Indications on the status of the signaling links at its signaling point, received from the MTP2s.
2. SNM messages received from the SNM functions at other signaling points.

Outputs. The actions of SNM result in the outputs listed below:

3. Commands to the MTP2s of the signaling links.
4. Commands to the SMH message routing function at its signaling point: for example, to divert signaling messages to a certain destination from their normal signaling link to an alternative link.
5. Indications (primitives) to the MTP-users at its signaling point, about the status of signaling route sets to individual destinations (Fig. 8.7-1). MTP-status, MTP-pause, and MTP-resume indicate that the signaling route set to a particular destination has become congested, unavailable, and available again.
6. Messages to SNMs at other signaling points.

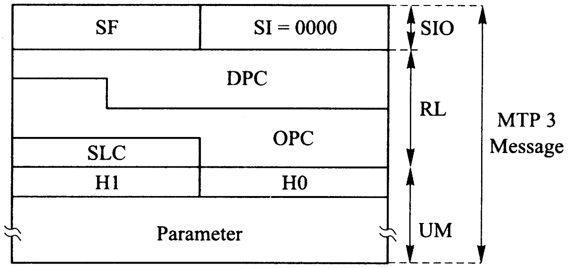


Figure 8.9-2. MTP3 message with a signaling network management message in the user message field. (From Rec. Q.704. Courtesy of ITU-T.)

8.9.2 SNM Messages

To send and receive messages to/from other signaling points, SNM uses the services of SMH in the same manner as the MTP-users.

The format of a MTP3 message with a SNM message in its user message (UM) field is shown in Fig. 8.9-2. The contents of routing label (RL) are shown in the ITU-T SS7 format.

SNM messages have service indicator code (SI) = 0000. RL contains OPC and DPC and, in lieu of the signaling link selector (SLS) parameter, the parameter *signaling link code* (SLC).

Signaling Link Code (SLC). Some SNM messages have status information about a particular signaling link. In these messages, the signaling link is identified by the combination of the parameters: DPC, OPC, and SLC. In SNM messages that do not pertain to a particular signaling link, the link code is set to SLC = 0.

Since the MTP3 messages that carry SNM messages do not have a SLS parameter, SMH message routing has to use special routing rules. When SLC = 0, the message can be routed on any link in the route set to the message destination. When SLC $\neq$ 0, it identifies a particular signaling link. In this case, the message has to be sent either on the identified link or on an alternate of the identified link, depending on the SNM message type. We will see the reason for this in the sections that follow.

Headings and Parameters. SNM messages always contain a heading and may include a parameter. The heading consists of the H0 and H1 fields. The code in H0 represents a group of functionally related messages, and the code in H1 indicates a particular message in the group—see Table 8.9-1. The table also lists the message acronyms and—if the message includes a parameter—the parameter name and acronym. In the sections that follow, the messages and parameters are denoted by these acronyms.

When a procedure involves signaling network management at several signaling points, we shall denote the SNM at a particular signaling point—say, point A—by SNM-A, and so on.

TABLE 8.9-1 Messages for Signaling Network Management

Message Name	Acronym	H0	H1	Parameter
Changeover Order	COO	0001	0001	FSNR (1)
Changeover Acknowledgment	COA	0001	0010	FSNR (1)
Changeback Declaration	CBD	0001	0101	CBC (2)
Changeback Acknowledgment	CBA	0001	0110	CBC (2)
Emergency Changeover	ECM	0010	0001	—
Emergency Changeover Acknowledgment	ECA	0010	0010	—
Transfer Prohibited	TFP	0100	0001	PAD (3)
Transfer Allowed	TFA	0100	0101	PAD (3)
Transfer Controlled	TFC	0011	0010	PAD (3)

Notes: (1) Forward sequence number of last accepted MSU. (2) Changeback code. (3) Point code of affected destination.

Source: Rec. Q.701. Courtesy of ITU-T.

8.9.3 Procedures for Congestion Control

The MTP2s of each signaling link constantly monitor the number of waiting messages in their output and retransmission buffers [11]. When this number exceeds a preset threshold value, SNM is alerted with an *onset of congestion* indication (Section 8.6.6). Outgoing signaling link congestion is determined independently at each end of the link. A signaling link between signaling points A and B can thus be congested at A and not congested at B.

When a signaling link is congested, SNM considers all destinations whose signaling route sets include that link as "congested." In Fig. 8.9-3, signaling end point A has two signaling links (SL_1 and SL_2), to the signal transfer points B and C in its region. The route set at A for all destinations thus consists of SL_1 and SL_2. When SNM-A receives an "onset of congestion" indication from the MTP2 of signaling link SL_1, it therefore concludes that its route sets to all destinations are congested.

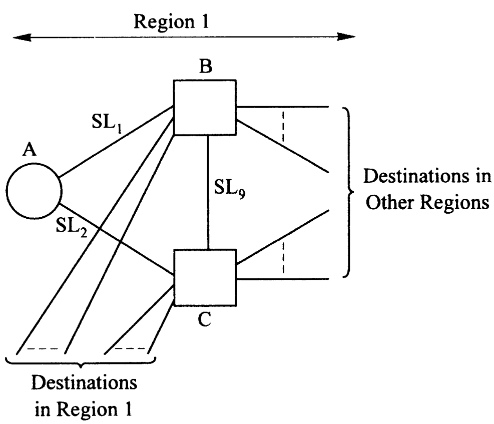


Figure 8.9-3. Congestion of SL_1 at signaling end point A.

It is important to keep congestion within limits, because otherwise the queuing delays for messages can become unacceptably large for call-control applications.

Congestion Control at Signaling End Points. In a signaling end point (SEP), the most likely cause of signaling link congestion is an excessive amount of messages originated by the MTP-users. SNM then passes *MTP-status* indications to all users, identifying the affected destination by its point code PAD (Fig. 8.7-1). In ITU-T SS7, separate indications have to be sent for each affected destination. In ANSI No. 7, certain point codes represent a group of destinations.

In response, the MTP-users reduce their outgoing message traffic to the affected destinations for a certain amount of time. For example, TUP and ISUP suspend the seizure of SS7 trunks to these destinations.

SNM keeps monitoring the outgoing messages to the affected destinations and, when necessary, repeats the MTP-status indications. When SNM receives an *end of congestion* indication from the MTP2 of the previously congested signaling link, it stops passing the MTP-status indications, and the MTP-users then end their restrictions on outgoing messages.

Congestion Control at Signal Transfer Points. When the SNM at a signal transfer point (STP) receives a congestion indication from one of its signaling links, it informs the signaling points from which it receives messages to the affected destinations, by sending Transfer Controlled (TFC) messages. For example, when SNM-D of signal transfer point D (Fig. 8.9-4) receives an onset of congestion indication from the MTP2 of signaling link SL_7, it considers its route to F as congested. Then, on receipt of the first message to F from a signaling end point, say, A, it sends a TFC message to SNM-A. Parameter PAD holds the point code of affected destination F. SNM-D repeats the TFC message each time it has received eight additional messages for F from A. The same procedure is carried out by SNM-D for all other SEPs from which it receives messages for F.

On receipt of each TFC message for affected destination F, the SNMs in the signaling end points that have originated messages to F pass MTP-status indications for affected destination F to their MTP-users.

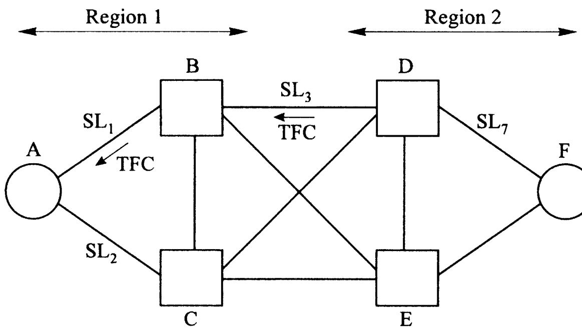


Figure 8.9-4. Congestion control by signal transfer point D.

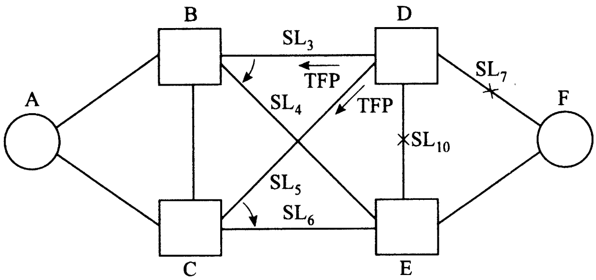


Figure 8.9-5. Rerouting of messages from A to F on failure of SL_7 and SL_{10}.

When SNM-D receives an indication that the congestion on SL_7 has ended, it stops sending TFC messages for affected destination F to the originating SEPs, which then end their restrictions on outgoing messages to F.

8.9.4 Rerouting of Messages

In Fig. 8.9-5, signaling links SL_7 and SL_{10} have failed simultaneously. Signal transfer point D is therefore no longer able to transfer messages to destination F.

SNM-D then alerts the SNMs in all directly connected signaling points (such as B and C) that it can no longer transfer messages to F, by sending Transfer Prohibited (TFP) messages. In these messages, PAD holds the point code of affected destination F.

We now examine the actions at signal transfer point B, in response to the TFP message. Suppose that SMH-B message routing normally routes messages for F in which SLS = XX0X and SLS = XX1X on, respectively, SL_3 and SL_4—see Fig. 8.8-3(*b*). After receipt of the TFP, SNM-B instructs SMH-B message routing to reroute messages to F with SLS = XX0X to link SL_4, starting with the messages to F that are waiting in the output buffer of SL_3. However, the messages waiting in the retransmission buffer of SL_3 are not retransmitted because SNM-B cannot determine whether they have reached their destination. The rerouting procedure thus may cause the loss of some messages.

Similar actions occur at signal transfer point C, and at all other signaling points that have received the TFP message.

When SL_7 and/or SL_{10} become available again, SNM-D informs the previously alerted signaling points with Transfer Allowed (TFA) messages, indicating that it can again transfer messages with destination F. These SPs then resume their normal routings of messages to that destination.

8.9.5 Signaling Route Set Unavailable

Certain combinations of multiple failures in a signaling network can completely disable a route set. Consider two signaling end points, say, A and F. When SNM-A determines that its route set to a destination F has become unavailable, it informs its MTP-users by passing *MTP-pause* indications (Fig. 8.7-1), which include the point code (PAD) of affected destination F. The users then take the necessary measures. For instance, if signaling point A has a TUP trunk group to F, TUP-A has to take this group out of service.

When SNM-A determines that the route set is no longer disabled, it informs its users with *MTP-resume* indications, again identifying the affected destination F by its point code.

8.9.6 Changeover and Changeback

When a signaling link fails or is taken out of service for maintenance, it becomes unavailable for the transfer of messages. The messages normally sent out on the link have to be diverted to an alternative signaling link. This requires changeover actions by the SNMs at both ends of the link. When the link becomes available again, changeback actions by the SNMs terminate these diversions.

We examine these procedures for the case that link SL_1 in Fig. 8.9-6 has failed. Signaling end point A has to divert all messages that are normally sent on SL_1 to SL_2, and signal transfer point B, which normally sends all messages with destination A on SL_1, has to divert these messages to SL_9, and the messages arrive at A on signaling link SL_2.

Whenever possible, the changeovers and changebacks are executed in such a way that the affected messages are still delivered in-sequence and without loss or duplication.

Changeover. When SNM-A decides that SL_1 has to be made *unavailable* for messages, it starts the changeover by ordering the MTP2-A of SL_1 to suspend sending messages and to "freeze" the contents of its output and retransmission buffers.

SNM-A then sends a Changeover Order (COO) message to SNM-B. The message identifies the unavailable link SL_1 by the combination of the values of parameters DPC, OPC, and SLC. Also included in the message is FSNR, the forward sequence number of the last message received on SL_1 that has been accepted at A (see Table 8.9-1).

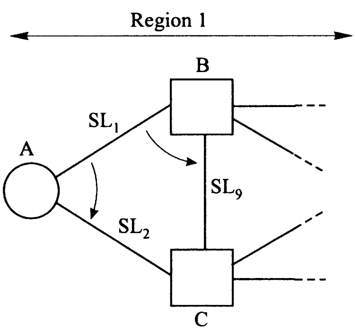


Figure 8.9-6. Changeover of signaling link SL_1.

For COO messages, SMH message routing selects an outgoing signaling link *other than* the link identified by SLC. In this example, the COO thus goes out on SL_2 and reaches B on SL_9.

When SNM-B receives the changeover order, it orders the MTP2-B of SL_1 to stop sending messages on SL_1 SNM-B then retrieves all messages not yet acknowledged by MTP2-A (i.e., messages with forward sequence numbers exceeding FSNR) from the retransmission buffer of SL_1 and moves them into the output buffer of alternative link SL_9. These messages are therefore sent out on that link.

In the next step, SNM-B retrieves the messages in the output buffer of SL_1 and moves them into the output buffer of SL_9. These messages are also sent out on that link. SNM-B then orders SMH-B message routing to route all subsequent messages that would normally be sent on SL_1, on SL_9.

As a result, all messages from B that are affected by the changeover are sent, in-sequence and without loss or duplication, on SL_9. They reach A on SL_2.

SNM-B now sends a Changeover Acknowledgment (COA) message to SNM-A, in which the parameter FSNR represents the forward sequence number of the last message it has accepted from SL_1. The COA message is also sent on SL_9 and reaches signaling point A on SL_2.

When SNM-A receives the COA message, it executes a similar procedure, which has the result that all messages waiting in the buffers of SL_1 are sent out on SL_2, without loss or duplication, and in-sequence. They reach signaling point B on SL_9.

Emergency Changeover. The above "best case" procedure cannot always be carried out, because failures in the MTP2-A of SL_1 may make the determination of FSNR impossible. In this case SNM-A sends an Emergency Changeover message (ECM), which does not include FSNR. When SNM-B receives the ECM, it ignores the contents of the retransmission buffer of SL_1 and only moves the contents of the output buffer to SL_9. This results in the loss of some messages.

A failure of SL_1 is likely to be noticed by both SNM-A and SNM-B. If SNM-A receives the changeover order from SNM-B after it has already started its changeover actions, it still acknowledges the order.

Changeback. When SL_1 becomes available again for the transfer of messages, it is necessary to end the diversion. This requires changeback actions by SNM-A and SNM-B.

SNM-A and SNM-B are informed by their respective MTP2s when SL_1 is again in working condition. Both SNMs may initiate the changeback. The procedure initiated by SNM-A is described below—the procedure at SNM-B is similar. The objective of the changeback procedures is to start using SL_1 again, preferably without causing message loss, duplication, or out-of-sequence delivery. The procedure is executed in two steps:

Step 1. SNM-A informs SMH-A message routing that messages should no longer be diverted to SL_2, but stored in a *changeback buffer* (CBB) instead. SNM-A then sends a Changeback Declaration (CBD) to SNM-B, identifying SL_1 by the combination of the values in DPC, OPC, and SLC. Changeback declarations, like changeover messages, are routed on the alternate of the identified link SL_1. In our example, the CBD goes out on SL_2. The message reaches signaling point B on SL_9, and SNM-B acknowledges with a Changeback Acknowledgment (CBA).

Step 2. When SNM-A receives the CBA, it knows that all diverted (to SL_2) messages that were on their way to B when it sent the CBD have been accepted there, because signaling links SL_2 and SL_9 accept incoming messages *in-sequence*. Therefore, the messages that have accumulated in the CBB can now be sent out without causing out-of-sequence (premature) deliveries at their destinations. SNM-A thus orders SMH-A message routing to send out the messages in changeback buffer CBB on SL_1, and then to resume using that link for subsequent messages.

CBD and CBA messages include a parameter called *changeback code* (CBC) (see Table 8.9-1). This parameter is needed because a SNM may be involved in several changebacks simultaneously. The parameter value identifies a particular changeback action—and the associated changeback buffer (CBB)—at signaling point A. When SNM-B acknowledges the changeback declaration, it copies the value of CBC in its acknowledgment. This enables SNM-A to select the proper CBB and signaling link for step 2.

8.10 ACRONYMS

ANSI	American National Standards Institute
BIB	Backward indicator bit
BSN	Backward sequence number
CB	Check bit
CBA	Changeback Acknowledgment message
CBB	Changeback buffer
CBC	Changeback code
CBD	Changeback Declaration message
COA	Changeover Acknowledgment message
COO	Changeover Order message
DPC	Destination point code
ECA	Emergency Changeover Acknowledgment message
ECM	Emergency Changeover message
ETSI	European Telecommunications Standards Institute
FIB	Forward indicator bit
FISU	Fill-in signal unit
FSN	Forward sequence number
FSNR	Forward sequence number of the most recently accepted message
H0	Message heading field 0

H1	Message heading field 1
IB	Input buffer
IP	Incoming processing
ISC	International switching center
ISUP	ISDN user part of SS7
ITU	International Telecommunication Union
ITU-T	Telecommunication Standardization Sector of ITU
kb/s	Kilobits per second
LC	Link control
LI	Length indicator
LSI	Link status information
MSU	Message signal unit
MTP	Message Transfer Part
MTP1	Message Transfer Part level 1
MTP2	Message Transfer Part level 2
MTP3	Message Transfer Part level 3
OB	Output buffer
OP	Outgoing processing
OPC	Originating point code
PAD	Point code of affected destination
PCM	Pulse code modulation
PCR	Preventive cyclic retransmission
RB	Retransmission buffer
RL	Routing label
RST	Route-set test message
SCCP	Signaling Connection Control Part
SDL	Signaling data link
SEP	Signaling end point
SI	Service indicator
SIB	Signaling link congested (LSSU)
SIE	Emergency alignment (LSSU)
SIF	Signaling information field
SIN	Normal alignment (LSSU)
SIO	Out of alignment (LSSU)
SIO	Service information octet
SIOS	Link out of service (LSSU)
SIPO	Processor outage (LSSU)
SL	Signaling link
SLC	Signaling link code
SLS	Signaling link selector
SMH	Signaling message handling
SNM	Signaling network management
SP	Signaling point
SSF	Subservice field
SS6	Signaling System No. 6

SS7	Signaling System No. 7
STEP	Signal transfer and end point
STP	Signal transfer point
SU	Signal unit
TFA	Transfer Allowed message
TFC	Transfer Controlled message
TFP	Transfer Prohibited message
TUP	Telephone User Part
UM	User message

8.11 REFERENCES

1. A. R. Modaressi and R. A. Skoog, Signaling System No. 7: A Tutorial, *IEEE Commun. Mag.*, **28**, No. 7, July 1990.
2. R. Manterfield, *Common Channel Signalling*, Chapter 5, Peter Peregrinus Ltd., London, 1991.
3. B. Law and C. A. Wadsworth, CCITT Signalling System No. 7: Message Transfer Part, *Br. Telecommun. Eng.*, **7**, Apr. 1988.
4. *Specifications of Signalling System No. 7*, Rec. Q.701, ITU-T Geneva, 1993.
5. *American National Standards for Telecommunications—Signaling System No. 7 (SS7)—Message Transfer Part*, ANSI T.111, New York, 1992.
6. *Specifications of Signaling System No. 7*, Rec. Q.702, ITU-T, Geneva, 1988.
7. *Specifications of Signaling System No. 7*, Rec. Q.703, ITU-T, Geneva, 1996.
8. *Specifications of Signaling System No. 7*, Rec. Q.706, ITU-T, Geneva, 1993.
9. *Specifications of Signaling System No. 7*, Rec. Q.704, ITU-T, Geneva, 1996.
10. *Specifications of Signaling System No. 7*, Annex A to Rec. Q.705, ITU-T, Geneva, 1993.
11. D. R. Manfield, G. Millsteed, and M. Zukerman, Congestion Controls in SS7 Signaling Networks, *IEEE Commun. Mag.*, **31**, No. 6, June 1993.

9

TELEPHONE USER PART

The Telephone User Part (TUP) was the first SS7 user part defined by ITU-T. An early version appeared in the 1980 CCITT *Yellow Book*. The 1985 *Red Book* and the 1989 *Blue Book* included a number of additions and modifications [1].

The term "telephone" dates back to the beginnings of the TUP development, when all calls were "speech" calls. The calls in the present telecommunication networks can also be used for facsimile and other data communications. TUP is primarily a link-by-link signaling system. It can be used on analog FDM trunks and on 64-kb/s digital trunks (see Sections 1.4 and 1.5).

The ITU-T specifications cover TUP applications in national networks and in the international network. TUP has been designed to be backward compatible with R2 and SS6 signaling (Chapters 4 and 5) and includes all features of these systems. National versions of TUP coexist with R2 signaling (Chapter 4) in the networks of several countries. Since R2 provides end-to-end signaling (between the originating and terminating exchanges of a connection), TUP includes a similar procedure.

One important aspect of TUP is its support of *digital connectivity* (the provision of transparent end-to-end 64-kb/s digital connections). The demand for this service is growing steadily, because of new digital customer equipment, such as high-speed facsimile machines. Telecoms in several countries have installed subscriber lines that can transfer 64-kb/s digital information. These lines are the precursors of the digital subscriber lines for the *integrated digital services network* (ISDN), which are discussed in Chapter 10.

TUP was introduced in the international network in the mid-1980s. It is also used—with some modifications—in the national networks of several countries. Telecoms in other countries have chosen to bypass TUP signaling. For example, the telecoms in the United States have gone directly from common-channel interoffice signaling (CCIS—see Chapter 5) to the ANSI-defined version of the ISDN user

Signaling in Telecommunication Networks, Second Edition, by John G. van Bosse and Fabrizio U. Devetak

part (ISUP). In Japan, a national version of ISUP has replaced multifrequency signaling.

Sections 9.1 through 9.5 primarily describe TUP as specified in ITU-T recommendations. Section 9.6 briefly discusses some aspects of national versions.

9.1 MESSAGES AND PRIMITIVES

This section outlines the structure of TUP messages and reviews the primitives that involve TUP.

9.1.1 General Message Structure

The general TUP message format is shown in Fig. 9.1-1 [2,3]. Octet (a) is the *service information octet* (SIO) (Section 8.8.3), consisting of *service indicator* (SI) and *subservice field* (SSF). The value SI = 0100 indicates a TUP message.

The *routing label* (RL) is in octets b through e and contains the *originating* and *destination point codes* (OPC, DPC) and the *signaling link selector* (SLS) (Section 8.8.2). The trunk for which the message is intended is identified by the combination of OPC, DPC, and CIC (circuit identification code).

9.1.2 Signaling Link Selector and Circuit Identification Code

Bits 8 to 5 of octet e have a dual function: they represent both the SLS and the four low-order bits (CIC_L) of the circuit identification code. The high-order bits (CIC_H) are in octet f. The value of SLS thus equals the value of CIC, modulo 16.

As a result, each trunk has an associated signaling route. This is one of the requirements for in-sequence message delivery (Section 8.8.5).

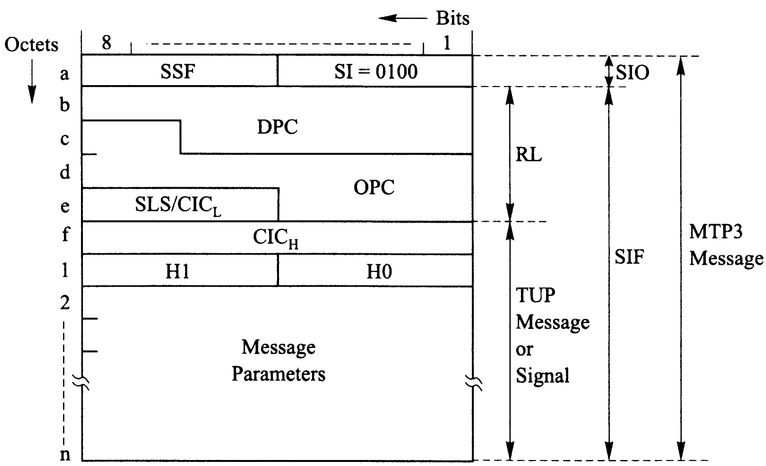


Figure 9.1-1. General format of TUP messages and signals. SIO, Service information octet; SIF, signaling information field.

9.1.3 TUP Messages and Signals

TUP literature makes a distinction between messages and signals. A TUP *signal* consists of octets f and 1 in Fig. 9.1-1. In a TUP *message*, one or more octets with message parameters follow octet 1.

9.1.4 Heading

The H1/H0 octet is known as the *heading* and identifies a particular message or signal. H0 represents a group of functionally related messages/signals, and H1 identifies a particular message/signal within that group. This structure is the same as the heading structure of signaling network management messages (Section 8.9.2).

9.1.5 Primitives

TUP communicates with the Message Transfer Part (MTP) with the primitives shown in Fig. 8.7-1. MTP-transfer requests and indications pass information in the SIO and SIF fields of TUP messages/signals from TUP to MTP, and vice versa.

The MTP-status, MTP-pause, and MTP-resume indications alert TUP that the signaling route set to a destination is congested, has become unavailable, and has become available again. These primitives contain the parameter PAD (point code of the affected destination).

9.2 CALL-CONTROL MESSAGES AND SIGNALS

This section describes the most important messages, parameters, and signals for TUP call control.

The octets shown in the figures of this section correspond to octets 1 through *n* of Fig. 9.1-1.

9.2.1 Initial Address Message

The Initial Address message (IAM) is the first forward message in a call setup. It contains the calling party category, a number of message indicators, and digits of the called party number—see Fig. 9.2-1.

Calling Party Category (CPC). This parameter is coded as follows:

Bits 654321	
000000	Unknown calling party category
000001	French-speaking operator
000010	English-speaking operator
000011	German-speaking operator
000100	Russian-speaking operator
000101	Spanish-speaking operator
001010	Ordinary calling subscriber
001101	Test call

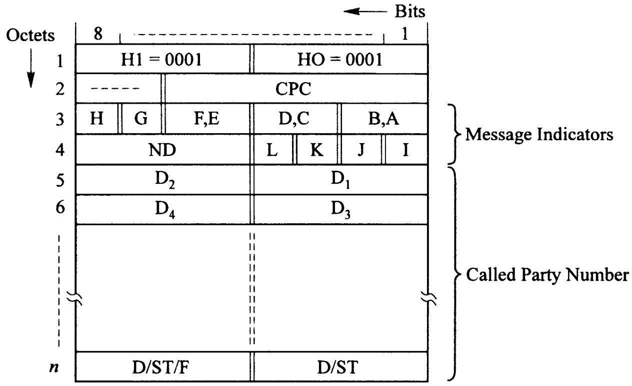


Figure 9.2-1. Initial Address message. (From Rec. Q.723. Courtesy of ITU-T.)

Message Indicators (A,...,K). A set of indicators is used by the exchanges along the connection during the call setup:

(BA)	Nature of (Called) Address Indicator
00	Subscriber number
10	National number (area code + subscriber number)
11	International number (country code + national number)

(DC)	Nature of Circuit Indicator
00	No satellite circuit in connection
01	One satellite circuit in connection

(FE)	Continuity Check Indicator
00	Continuity check not required
01	Continuity check required on this circuit
10	Check not required on this circuit but has been performed on a previous circuit

(G)	Echo Control Indicator
0	Outgoing suppressor/canceler not included
1	Outgoing suppressor/canceler included

(H)	Incoming Call Indicator
0	Not an incoming international call
1	Incoming international call

(I)	Redirected (Forwarded) Call Indicator
0	Not a redirected call
1	Redirected call

(J)	Digital Path Indicator
0	All digital (64-kb/s) path not required
1	All digital (64-kb/s) path required

(K)	Signaling Indicator
0	Path with SS7 signaling not required
1	Path with SS7 signaling required

Number of Digits (ND). This indicates the number of digits of the called address that are included in the IAM.

Digits (Di). Digit values 0 through 9 are coded 0000 through 1001; codes 11 and 12 are coded as 1011 and 1100, and *end of address* digit (ST) is coded as 1111.

SS7 uses 0000 as "digit 0" and as a "filler" code. If 0000 appears in bits 8 to 5 of octet *n*, it means "digit 0" if ND is *even* and "filler" if ND is *odd.*

9.2.2 Initial Address Message with Additional Information (IAI)

This is an Initial Address message that includes one or more additional parameters. These parameters are *optional*: they are included only when necessary. Octets 2 through *n* in Fig. 9.2-2 are the same as in an IAM.

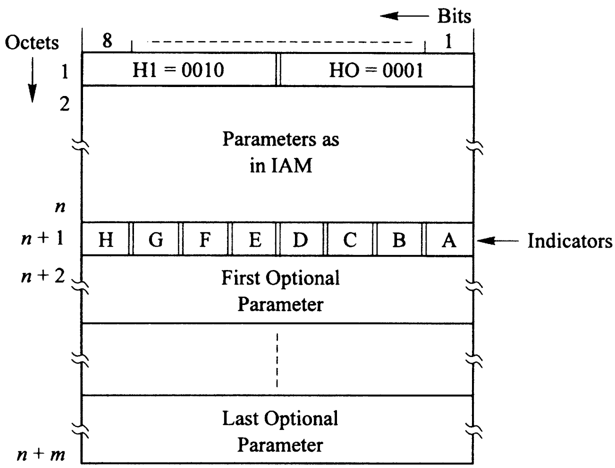


Figure 9.2-2. Initial Address Message with Additional Information (IAI). (From Rec. Q.723. Courtesy of ITU-T.)

Octet $n + 1$ contains indicator bits that show whether the optional parameters are present (1) or not (0). ITU-T has defined three optional parameters and indicator bits:

Indicator Bit	Optional Parameter
E	Calling line identity
H	Original called address
B	Closed user group information

When the IAI includes more than one optional parameter, they appear in the alphabetical order of their indicator bits.

Calling Line Identity and Original Called Address. The format of these parameters is shown in Fig. 9.2-3. ND indicates the number of digits; bits B and A indicate the nature of the address:

Bits BA	Nature of Address
00	Subscriber number
10	National number
11	International number

Indicator bit C is used in the calling line identity only. It indicates whether the calling party allows the presentation of his number to the called party:

Bit C	Calling Number Presentation
0	Allowed
1	Not allowed

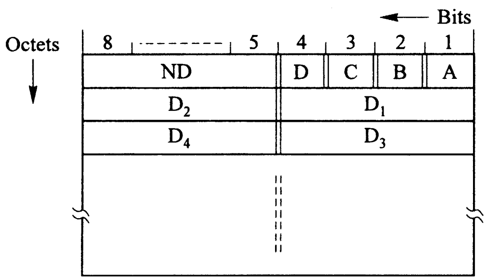


Figure 9.2-3. Format of calling line identity and original called address. (From Rec. Q.723. Courtesy of ITU-T.)

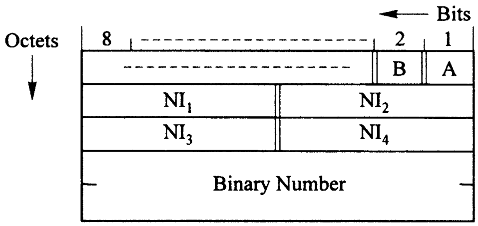


Figure 9.2-4. Coding of closed user group information. (From Rec. Q.763. Courtesy of ITU-T.)

Closed User Group Information. In some countries, subscriber lines and subscriber numbers can be marked at their local exchanges as belonging to a *closed user group* (CUG). Special call-control procedures apply to the members of these groups (Section 9.4).

The CUG information parameter is shown in Fig. 9.2-4. Bits B and A indicate whether a calling CUG member has "outgoing access" (is allowed to make calls to subscriber numbers that are not members of the caller's CUG):

Bits BA	Outgoing Access
10	Allowed
01	Not allowed

The second and later octets hold the *CUG interlock code*, which uniquely identifies a CUG worldwide. *Network identity* digits, $NI_1, \ldots, NI_4$, identify the network of a particular telecom, and the binary number identifies a CUG that is administered by this telecom.

9.2.3 Subsequent Address Messages

TUP address signaling allows both *en-bloc* signaling, in which the complete called number is included in the IAM or IAI, and *overlap* signaling, in which IAM (or IAI) includes only those digits that are needed for outgoing trunk selection at the next exchange.

Under overlap signaling, the later address digits are sent in one or more subsequent messages, of which there are two types.

The Subsequent Address message (SAM), identified by H1 = 0011, H0 = 0001, is an IAM (Fig. 9.2-1) without octets 2 and 3. The number of included digits is again indicated by the value of ND. Bits L, K, J, and I are set to zero.

The Subsequent One-Digit Address message (SAO) has the heading H1 = 0100, H0 = 0001, and one octet that contains one digit, in bits 4 to 1. Bits 8 to 5 are coded 0000.

9.2.4 Address Complete Message (ACM)

This is a backward message, originated by the last exchange in a connection of TUP trunks. It includes one octet with indicators—see Fig. 9.2-5:

(BA)	Call Charging Indicator
00	No instructions on charging
10	Charge
11	Do not charge

(C)	Called Subscriber Status
0	No information available
1	Subscriber free

(D)	Incoming Echo Control Indicator
0	Incoming suppressor/canceler not included
1	Incoming suppressor/canceler included

(E)	Call Forwarding Indicator
0	Call not forwarded
1	Call forwarded

(F)	Signaling Path Indicator
0	Not a completely SS7 signaling path
1	Completely SS7 signaling path

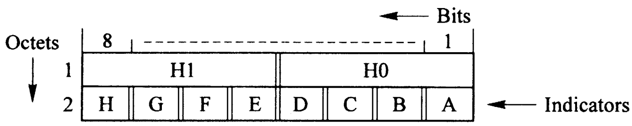


Figure 9.2-5. Address Complete message (ACM), Automatic Congestion Control message (ACC), and General Request message (GRQ). ACM: H1 = 0001, H0 = 0100. ACC: H1 = 0001, H0 = 1001. GRQ: H1 = 0001, H0 = 0011. (From Rec. Q.723. Courtesy of ITU-T.)

9.2.5 General Request Message (GRQ)

This backward message requests actions and/or information from a preceding exchange in the connection. The message format (Fig. 9.2-5) consists of one octet of indicators, each of which represents a request for a particular action or information item (parameter):

Indicator	Requested Action or Parameter
A	Calling party category
B	Calling line identity
C	Original called address
D	Malicious call identification
E	Call hold
F	Inclusion of outgoing echo controller

To request a particular action or information item, the corresponding indicator bit is set to 1.

9.2.6 General Forward Setup Information Message (GSM)

This forward message is sent in response to a received GRQ message. It includes an octet with indicator bits and can also contain one or more of the requested parameters—see Fig. 9.2-6. Each indicator bit corresponds to a particular requested action or parameter:

Indicator	Action or Parameter
A	Calling party category
B	Calling line identity
D	Original called address
E	Outgoing echo controller included
F	Malicious call identification
G	Call will be held under control of terminating exchange

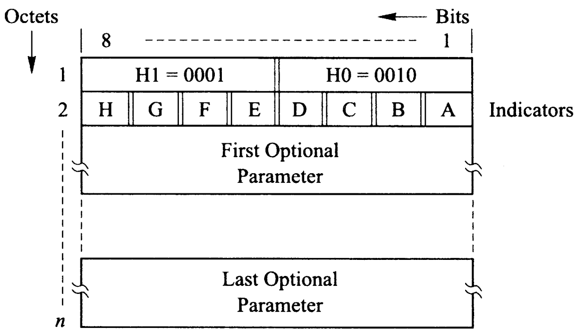


Figure 9.2-6. General Forward Setup Information message (GSM). (From Rec. Q.723. Courtesy of ITU-T.)

The value 1 of an indicator bit signifies that the requested action has been (or will be) taken, or that the requested parameter is included in the message. Included parameters appear in alphabetical order of the corresponding indicator bits.

The coding of the calling party category is as in the IAM (Section 9.2.1). The coding of the calling line identity and original called address is as in the IAI (Section 9.2.2).

9.2.7 Automatic Congestion Control Message (ACC)

This message is sent by an exchange whose control equipment is congested (overloaded) and requests the directly connected exchanges to reduce the number of seizures of outgoing trunks to the exchange—see Fig. 9.2-5.

Indicator bits BA indicate the congestion level:

Bits AB	Congestion Level
01	Level 1 (moderate congestion)
10	Level 2 (severe congestion)

9.2.8 Call-Control Signals

These signals consist of octets a through l of Fig. 9.1-1.

Continuity Signals. These forward signals report the success or failure of a continuity check on a trunk.

Signal	Meaning	H1	H0
COT	Continuity check success	0011	0010
CCF	Continuity check failure	0100	0010

Unsuccessful Backward Signals. These signals are sent backward by the exchange that determines that the call setup cannot be completed and indicate the

TABLE 9.2-1 Unsuccessful Backward Setup Signals (H0 = 0101)

H1	Name	Acronym
0001	Switching Equipment Congestion	SEC
0010	Circuit Group Congestion	CGC
0011	National Network Congestion	NNC
0100	Address Incomplete	ADI
0101	Call Failure	CFL
0110	(Called) Subscriber Busy	SSB
0111	Unallocated Number	UNN
1000	(Called Line) Out of Service	LOS
1001	Send Special Information Tone	SST
1010	Access Barred	ACB
1011	Digital Path not Provided	DPN

Source: Rec. Q.723. Courtesy of ITU-T.

TABLE 9.2-2 Call Supervision Signals (H0 = 0110)

H1	Name	Acronym
0000	Answer	ANU
0001	Answer, Charge	ANC
0010	Answer No Charge	ANN
0011	Clear-Back	CBK
0100	Clear-Forward	CLF
0101	Re-Answer	RAN
0110	Forward-Transfer	FOT

Source: Rec. Q.723. Courtesy of ITU-T.

reason for the setup failure (subscriber busy, address incomplete, etc.). The signal acronyms and heading codes are listed in Table 9.2-1.

Call Supervision Signals. This group consists of forward and backward signals that are sent on the occurrence of events in a connection that has been set up (answer, clear-forward, etc.). The signals and their heading codes are listed in Table 9.2-2.

9.3 BASIC SIGNALING SEQUENCES

This section illustrates the use of TUP call-control messages, parameters, and signals, with a number of basic signaling sequences [4]. References to the descriptions in Section 9.2 are included. It is helpful to look up these references while reading the examples that follow.

9.3.1 Successful National Call

Figure 9.3-1 shows the signaling for a successful call in a national network from subscriber S_1 to subscriber S_2. Trunks T_1 and T_2 have TUP signaling. The address signaling is *en-bloc*. In this example, continuity checks are made on both trunks.

Setup of Connection. Having received the called number from S_1, exchange V seizes trunk T_1 and sends and IAM, in which the message indicator bits (FE) are set to "continuity check required" (Section 9.2.1). It also connects a continuity-test transceiver to T_1.

When exchange W has received the IAM, it knows that a continuity check will be made and therefore connects the send and receive channels of T_1 to each other (loop-back). After analyzing the contents of IAM, exchange W seizes trunk T_2, sends an IAM that indicates "continuity check required," and attaches a continuity-test transceiver to T_2. On receipt of the IAM, exchange X establishes a loop-back on trunk T_2.

Assuming that the transmission of trunk T_1 is in working condition, the transceiver at exchange V receives the check tone. The exchange then disconnects the

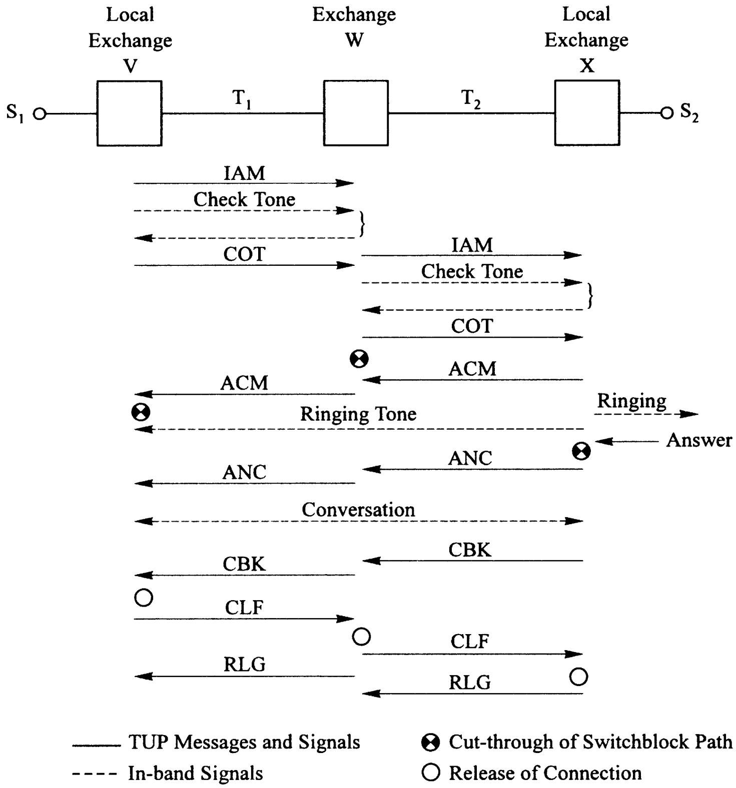


Figure 9.3-1. A successful national call.

transceiver from T_1 and sends a COT (continuity signal, Section 9.2.8). On receipt of COT, exchange W ends the loop-back on trunk T_1.

The continuity check of T_2 is also successful, and exchange W detects the check tone on T_2. Since W has already received the COT signal for T_1, it now completes the call setup, removing the transceiver from T_2, cutting through a switchblock path between T_1 and T_2, and sending a COT signal to exchange X. If W had received the check tone on T_2 but no COT signal for T_1, it would have waited for the COT signal before completing the setup.

Local exchange X examines the called number in the received IAM. We assume that the number is complete and valid, and that subscriber S_2 is free. From the IAM, exchange X knows that a continuity check will be made on T_2. It therefore awaits the COT signal for T_2 before proceeding: When the signal has been received, X sends an ACM (Address Complete message, Section 9.2.4), rings S_2, and sends ringing-tone on T_2. Exchange W repeats the ACM to exchange V, which then cuts through a switchblock path between S_1 and T_1. Subscriber S_1 now hears the ringing-tone.

When S_2 answers, exchange X cuts through a path between T_2 and S_2 and sends an ANC (Answer, Charge signal, Table 9.2-2). The signal is repeated to exchange V, which now starts to charge for the call. The conversation begins.

Message Indicator Settings. We briefly examine the settings of some IAM message indicators (Section 9.2.1) in the example. The call is national, and indicator BA is therefore set to "subscriber number" or "national number," and H is set to "not an international incoming call."

In Signaling System No. 6, exchanges made continuity checks on all outgoing trunks. In TUP signaling, the exchange decides whether to perform the check, based on stored information about the trunk. For example, continuity checking is not required for digital trunks, which are carried on a 24- or 30-channel time-division multiplexed transmission system. This is because the transmission system has no hardware units for individual trunks, and transmission problems can be detected at the multiplex port in the exchange. In this example, exchange V has set indicator FE to "continuity check required."

If V had decided not to make the check, it would have set FE to "continuity check not required." Exchange W then would not loop back trunk T_1 and would not wait for a COT signal before proceeding with the setup.

Assuming that the subscriber line of S_1 is an analog line, bit J indicates "all digital path not required," and exchange W can then select an analog or digital outgoing trunk. Also, on calls from analog subscribers, exchange V usually indicates (bit K) that SS7 signaling is not required, and exchange W could therefore have selected an outgoing trunk with another signaling type. The settings of other message indicators are discussed in later examples.

Indicator Settings in ACM. The ACM sent by exchange X, and repeated by exchange W, includes indicators with information for originating exchange V (Section 9.2.4). Bits BA indicate whether V has to charge S_1 for the call. Exchange X knows the status of called subscriber S_2 and sets indicator C to "subscriber free." The call has not been forwarded from S_2, and this is indicated by bit E. T_1 and T_2 have TUP signaling, which is indicated by bit F.

Release of Connection. In TUP signaling, the release of a connection takes place in the same way as in the signaling systems described earlier: it is initiated by originating exchange V. Connections are normally controlled by the calling party (Section 2.1.2). When the calling party disconnects, V immediately initiates the release. In Fig. 9.3-1, the called party disconnects first. Exchange X sends a CBK (Clear-Back) signal, which is repeated by exchange W. When exchange V receives the CBK signal, it starts a 30–60 s timer and initiates the release when the calling party disconnects, or when the timer expires, whichever occurs first. The release of trunk T_1 then takes place as follows. Exchange V disconnects the trunk at its end and sends a CLF (Clear-Forward) signal to exchange W. This exchange then clears T_1 at its end, and sends a RLG (Release-Guard) signal to indicate that the trunk is now available for new calls. Trunk T_2 is released in the same manner.

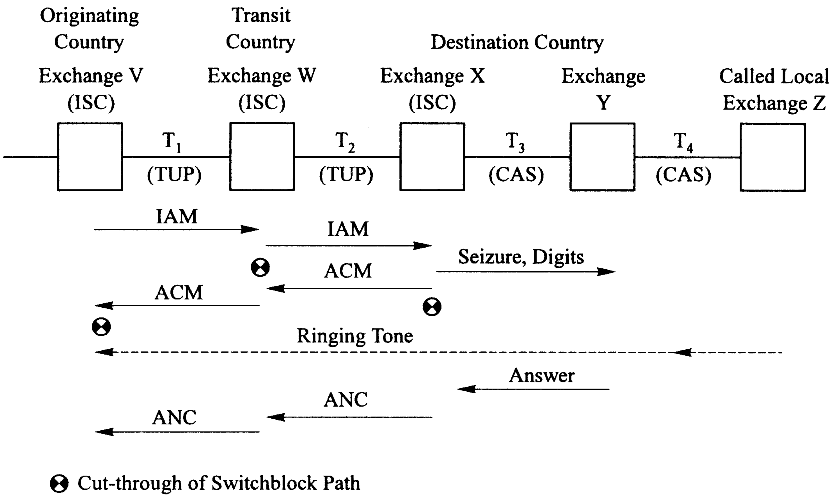


Figure 9.3-2. A successful international call.

9.3.2 Successful International Call

Figure 9.3-2 shows the setup of a successful international call. Exchanges V, W, and X are *international switching centers* (ISCs) in the originating country, a transit country, and the destination country. Exchanges V and W have seized international trunks (T_1, T_2) with TUP signaling. In this example, address signaling is *en-bloc*, and no continuity checks are made.

Exchange X seizes a national trunk T_3 with channel-associated signaling (CAS)—see Chapter 4. Exchange X is thus the last exchange in the TUP signaling segment of the connection. We examine the information in the IAM and ACM.

IAM Information. Exchange V has seized trunk T_1 to a transit country. The called number in its IAM (Section 9.2.1) is therefore an international number, and indicator (BA) is set to "international number." Exchange W seizes trunk T_2 to the destination country. It therefore removes the country code from the received called number, places the national called number in its IAM, and sets BA to "national number." Indicator H is set to "incoming international call" by the originating ISC (exchange V).

Some international trunk groups are carried on satellite transmission systems. It is desirable to have at most one satellite trunk in a connection (Section 1.4.6). If trunk T_1 is a satellite trunk, exchange V sets the IAM indicator DC to "satellite in connection," and exchange W then selects a terrestrial outgoing trunk.

Address Complete Message. Since trunk T_3 has CAS, the ACM is originated by exchange X. It informs exchanges W and V that they can discard the called number and other setup data. In this example, ACM indicator F (Section 9.2.4) is set to "not a completely SS7 signaling path."

On subscriber-dialed international calls, the exchanges in the originating country do not know whether a received called international number is complete, and the called number in the IAM sent by exchange V therefore does not include a ST (end of address digit).

Exchange X, the ISC in the destination country, has to determine whether the received national called number is complete, This is simple when the destination country has a uniform numbering plan (fixed-length national numbers), or when the country has variable-length national numbers and the received number has the maximum possible length. Otherwise, exchange X waits 4–6 s after the receipt of the IAM. If no subsequent address message (Section 9.2.2) is received during this interval, it sends the ACM.

Echo Control. The lengths of the international trunks may make echo control necessary. It is desirable to include only one pair of echo-control devices in a connection (Section 1.7.2). In TUP signaling, this is accomplished in the following manner. Suppose that exchange V includes an outgoing echo-control device on trunk T_1. It sets IAM indicator (G) to "outgoing echo control included." Exchange W repeats this indication in its IAM. If exchange X includes an incoming echo-control device on trunk T_2, it sets ACM indicator (D) to "incoming echo control included" (Section 9.2.4). In this case, exchange W merely repeats the ACM. However, if exchange X cannot include an incoming echo control on trunk T_2, it sets indicator (D) to "incoming echo control not included," and exchange W then includes an incoming echo-control device on trunk T_1.

9.3.3 Setup Problems

Dual Seizures. When a bothway trunk, say, trunk T_1 in Fig. 9.3-2, is seized simultaneously by exchanges V and W, the exchanges receive unexpected IAMs in response to their sent IAMs. In TUP signaling, the exchange with the higher point code (PC) controls the trunks with even CICs (circuit identification codes), and the exchange with the lower PC controls the trunks with odd CICs.

On a dual seizure, the controlling exchange continues its call setup, and the noncontrolling exchange "backs off" and attempts to seize another trunk for its call.

Response to Unsuccessful Backward Setup Signals. An exchange unable to extend a call setup, informs the preceding exchange with one of the unsuccessful backward setup signals listed in Table 9.2-1.

We first examine a national connection with SS7 signaling all the way (Fig. 9.3-3). When local exchange X determines that called subscriber S_2 is busy, it sends a Subscriber Busy (SSB) signal to exchange W. This exchange then initiates the release of trunk T_2, sends a Clear-Forward (CLF) signal, and also repeats the SSB signal. Exchange V then initiates the release of trunk T_1 and connects a busy-tone source to the line of S_1. Other problems detected at exchange X (called number incomplete or not allocated, called line out of service, etc.) are handled in the same manner.

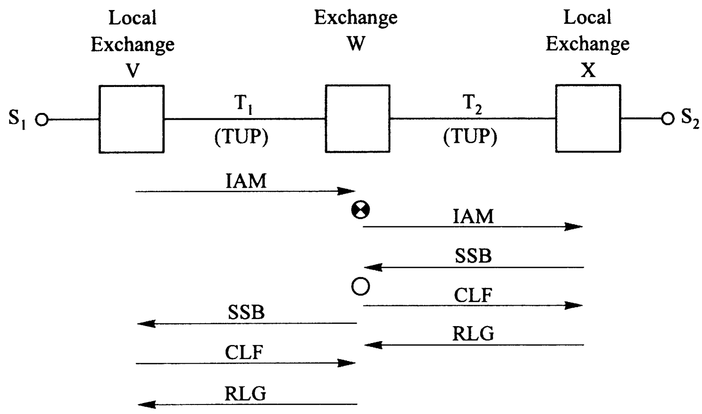


Figure 9.3-3. Release of a failed connection. Signaling path is TUP all the way.

In Fig. 9.3-2, trunks T_1 and T_2 have TUP signaling, but the signaling on T_3 and T_4 is channel-associated. Suppose now that the setup fails at the called local exchange Z. If T_3 and T_4 have national R2 signaling (Section 4.6), Z uses a group-B signal to inform exchange X that the called subscriber is busy, or that the called number is not allocated. Exchange X then releases T_3 and sends a SSB or Unallocated Number (UNN) signal for T_2, and exchange W initiates the release of the trunk. If trunks T_3 and T_4 have a signaling system that cannot inform exchange X about setup problems, local exchange Z connects its incoming trunk to a busy-tone, reorder-tone, or a recorded announcement. The calling subscriber then disconnects, and this initiates the release of the connection.

Automatic Rerouting. Exchanges can be programmed to perform automatic rerouting (Section 1.3.4). Suppose that in Fig. 9.3-2 international exchange V has seized trunk T_1 and receives a Switching Equipment Congestion (SEC) or Circuit Group Congestion (CGC) signal from exchange W. If V is arranged for automatic rerouting, it releases T_1 and attempts to setup the call on a trunk in a route to the destination that does no pass through exchange W.

Timeouts. TUP call setup procedures include a number of timeouts that cause a call setup to be abandoned. For example, when an exchange does not receive an ACM within 20–30 s after sending the final address message for the call, or does not receive an Answer signal within 2–4 minutes after the receipt of the ACM, it clears the forward connection and sends a Call Failure (CFL) signal to the preceding exchange. Also, when an exchange receives an IAM with message indicator bits (FE) indicating that the preceding exchange will make a continuity check, it returns a CFL signal if it does not receive the COT signal within 10–15 s.

This concludes the overview of basic TUP signaling for call control.

9.4 TUP SUPPORT OF ADDITIONAL SERVICES

This section describes a number of services that do not fall into the category of "basic" call control and the manner in which these services are supported in TUP signaling [4,5].

9.4.1 Services

The services to be described fall into one of the following classes.

Malicious Call Identification. Multifrequency Compelled (MFC) signaling in many national networks includes procedures to identify malicious callers (Section 4.6). These are also included in TUP signaling.

Digital Connectivity. Telecoms in France [8] and several other countries have introduced digital subscriber lines, without waiting for the completion of the specifications for digital subscriber lines in the integrated services digital network (ISDN).

Normal subscriber signaling (Chapter 3) is used on these lines. When the connection has been set up, the subscribers communicate with 64-kb/s digital data.

Supplementary Services. ITU-T has defined a number of supplementary services. These services require connections with SS7 (TUP or ISUP) signaling all the way.

Calling Line Identification. Subscribers marked for this service at their local exchanges have devices attached to their telephones that display the calling party number. In the United States this service is known as "caller ID" (Section 3.7.2).

Call Forwarding Service. In this service, a subscriber can indicate to his local exchange that, until further notice, incoming calls should be forwarded to another subscriber.

Closed User Group Service [5]. In this service (not offered in the United States), a multilocation business customer can register its lines as members of a closed user group (CUG). A CUG is identified worldwide by four digits that represent a particular telecom and an interlock code assigned by the telecom to the CUG (Section 9.2.2). A CUG restricts the calls that can be made and received by its members. A CUG has three characteristics that can be specified by the business customer:

1. *Outgoing Access*. Members of a CUG with no outgoing access (NOA) can only make calls to other members of their CUG. Members of a CUG with outgoing access (OA) have no restrictions on their outgoing calls.

2. *Incoming Access*. Members of a CUG with no incoming access (NIA) can receive calls from other members of their CUG only. Members of a CUG with incoming access (IA) have no restrictions on their incoming calls, except as noted below.
3. *Incoming Calls Barred*. Members of a CUG with incoming calls barred (ICB) are not allowed to receive calls from other members of their CUG. Members of a group without incoming calls barred (NICB) can receive calls from other members of their CUG.

The decision to allow or to reject a call involving members of a CUG is made by the local exchange of the called party (see Section 9.4.7).

9.4.2 The GRQ-GSM Procedure

This procedure, which is used in some of the services outlined earlier, is the functional equivalent of end-to-end interregister signaling in the R2 signaling systems (Sections 4.4.4 and 4.4.5). It allows the called exchange to make a request to the calling exchange, during the setup of the call. In Fig. 9.4-1, called exchange X has received an IAM and sends a General Request message (GRQ). Each message indicator bit represents a particular request from the called exchange (Section 9.2.5). When a request is made, the corresponding indicator bit is set to 1.

The requests are of two types: requests for additional information (calling line identity, original called address, etc.) and requests for special call-control actions (malicious call identification, holding the call under control of the called party, etc.). When a particular item is requested, the corresponding indicator is set to 1.

The GRQ is transferred backward, along the exchanges in the connection, and reaches calling exchange V. This exchange responds with a General Forward Setup Information message (GSM). Message indicator bits (Section 9.2.6) confirm (1) or deny (0) the individual requests. The parameters included in the message appear in the alphabetical order of their indicator bits. The GSM is transferred forward and reaches exchange X.

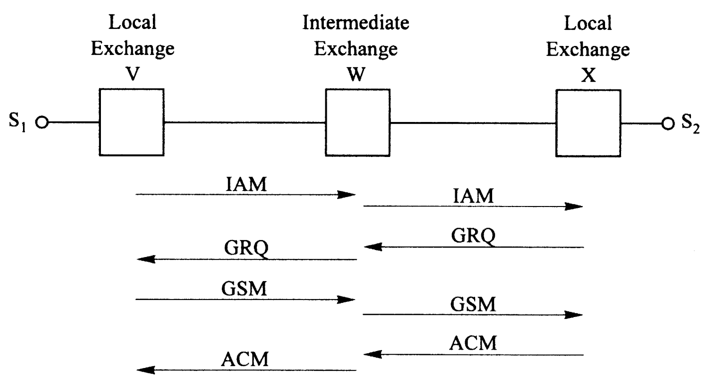


Figure 9.4-1. Request for information or action, and response.

A GRQ-GSM sequence has to be completed before the called exchange sends its ACM (address complete message) for the call.

9.4.3 TUP Support of Malicious Call Identification

The directory number of a subscriber who is subjected to malicious calls can be marked in his local exchange. Suppose that the directory number of a subscriber S is marked in this way.

When the local exchange of S receives an IAM or IAI for a call to S, it has to obtain the number of the calling party. If the initial address message is an IAI, this number may already be included. Otherwise, the exchange requests the calling number with a GRQ (Section 9.2.5).

If the call turns out to be malicious, subscriber S alerts his exchange with a flash signal (Section 3.3.2). The exchange then prints a record that includes the date and time-of-day of the call, and the calling and called numbers.

In some countries, the connections of malicious calls are also held under control of the called exchange, for tracing purposes. In these countries, the local exchange that receives an IAM or IAI for a call to subscriber S also requests "call holding." If the exchange receives a flash signal from S, it does not send a CBK (Clear-Back) signal when subscriber S disconnects. Moreover, when the calling party disconnects, the calling exchange will hold the connection until it receives a CBK signal from the called exchange. This enables personnel at the calling and called exchanges to trace the connection.

9.4.4 TUP Support of Digital Connectivity

For this service, the path between the calling and called digital subscriber lines has to be completely digital (64 kb/s in both directions). This requires that all exchanges in the connection have digital switchblocks, and that all trunks are digital trunks. These connections also require TUP signaling all the way. When a call originates on a (pre-ISDN) digital subscriber line, the originating exchange selects a digital outgoing TUP trunk to a next exchange that has a digital switchblock and informs the succeeding exchanges that the call is digital, by setting the IAM message indicators J and K to "digital path required" and "path with SS7 signaling required" (Section 9.2.1).

9.4.5 TUP Support of Calling Line Identification

This service requires the transfer of the calling line identity (calling party number) across the network. The service can be provided in two ways.

In some countries, the first address message of a call is always an IAI (Initial Address Message with Additional Information) in which the calling number is included (Section 9.2.2).

In other countries, the first message of a call is an IAM. The called exchange checks whether the called party is registered for the calling line presentation service. If so, the called exchange obtains the calling line identity with the GRO-GSM procedure.

The format of the calling line identity parameter is shown in Fig. 9.2-3. Subscribers can be marked at their local exchanges as allowing, or not allowing, the "presentation" of their number to called subscribers, and the originating exchange sets indicator bit C of the calling line identity parameter to "presentation allowed" or "presentation not allowed." The called exchange delivers the calling number only when the presentation is allowed.

9.4.6 TUP Support of Call Forwarding

A subscriber, say, S_1, who is registered at an exchange for call forwarding can activate and deactivate the service by dialing a service access code (Section 3.7.1). On activation, S_1 also dials the number of subscriber S_2, to whom incoming calls should be forwarded.

When a call for S_1 arrives while the service is active, her local exchange extends the call setup to S_2 and sets indicator bit I in its outgoing IAM to "forwarded call" (9.2.1).

If the exchange serving S_2 receives the IAM of the forwarded call and determines that S_2 also has call forwarding and has activated his service, it does not forward the call again, but abandons the setup. The call is thus forwarded only once. This avoids circular routing problems, for example, when S_1 is forwarding calls to S_2, and S_2 is forwarding calls to S1.

9.4.7 TUP Support of Closed User Groups

A local exchange has stored data that enable it to determine whether an attached subscriber line (or the directory number of the line) is a CUG member, and to obtain the interlock code and the characteristics of the CUG.

The decision to allow or reject a call involving one or two CUGs is made by the called exchange and is based on the calling and/or called CUG. Information on the called CUG is stored at the called exchange. Information on the calling CUG is stored at the originating exchange. The called exchange knows whether the calling party is a CUG member because, on calls made by CUG members, the initial message sent by the originating exchange is an IAI in which indicator B of octet $n+2$ in Fig. 9.2-2 is set to "CUG information included" and includes the CUG information (Section 9.2.2).

The criteria for accepting or rejecting the incoming call are listed in Table 9.4-1. If the calling and called parties are both CUG members, the decision depends on whether the members belong to the same CUG (indicated as CUG match in the table), and on the incoming and outgoing characteristics of the CUG (or CUGs).

If only the called party is a CUG member, the call is allowed if the called CUG has incoming access (IA). If only the calling party is a CUG member, the call is allowed if the calling CUG has outgoing access (OA).

TABLE 9.4-1 Acceptance and Rejection of Calls Involving Closed User Groups (CUGs)

		Called CUG Type				
Calling CUG Type	CUG Match	NIA-NICB	NIA-ICB	IA-NICB	IA-ICB	Non-CUG
NOA	Y	A	R	A	R	R
	N	R	R	R	R	
OA	Y	A	R	A	R	A
	N	R	R	A	A	
Non-CUG	—	R	R	A	A	A

Note: NOA, No outgoing access; OA, outgoing access; Non-CUG, calling or called party not a member of a CUG; NIA-NICB, no incoming access, incoming calls not barred; NIA-ICB, no incoming access, incoming calls barred; A, call accepted by called local exchange; R, call rejected by called local exchange; Y, yes; N, No.

Source: Rec. Q.723. Courtesy of ITU-T.

9.5 OTHER TUP PROCEDURES, MESSAGES, AND SIGNALS

This section describes a number of TUP procedures, messages, and signals that are not directly related to the control of calls [4,6]. We consider a trunk group (TG) between exchanges V and W—Fig. 9.5-1.

9.5.1 Overload at Exchange

During moments of high activity, the call processing equipment at an exchange may become overloaded. Suppose that this is the case for exchange W. The exchange then informs exchange V (and all other exchanges to which it is connected by a TUP trunk group) by sending Automatic Congestion Control (ACC) messages (Section 9.2.7). The ACC messages to exchange V are sent when exchange W receives a CLF signal for a trunk in group TG. Each ACC message is followed by a RLG signal, which is the normal response to the CLF signal.

ACC messages include one octet with indicators (Fig. 9.2-5). Bits B and A indicate the level of congestion:

(BA)	Meaning
01	Level 1 congestion (moderate)
10	Level 2 congestion (severe)

In response to the ACC message, exchange V reduces the number of seizures of trunks in trunk group TG. The degree of the reduction depends on the congestion level of exchange W. For example, exchange V may be programmed to abort

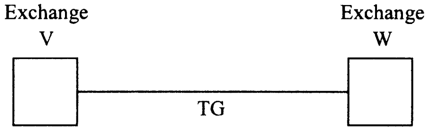


Figure 9.5-1. Exchanges V and W, interconnected by trunk group (TG).

every fourth seizure of a trunk in the group (25% reduction) for level 1 congestion, and all seizures (100% reduction) for level 2 congestion. If exchange V has other route choices for the call, it tries to setup the call on one of these routes. Otherwise, it abandons the setup.

Exchange V keeps reducing the number of seizures of trunks in group TG until some 5 s have elapsed without the receipt of a new ACC message from exchange W.

9.5.2 Trunk Blocking and Unblocking

Maintenance personnel (or an automatic process) in exchanges V and W can block a trunk. Suppose that exchange W needs to test a trunk T in group TG. It then sends a BLO (Blocking) signal for the trunk to exchange V—see Table 9.5-1. The latter exchange acknowledges with a BLA (Blocking Acknowledgment) signal. The affected trunk T is identified by DPC, OPC, and CIC (Fig. 9.1-1). If trunk T is carrying a call, the call is not interrupted, and exchange W starts the test only after the call has ended. Exchange V will not seize T for new calls, but will process test calls from exchange W. In test calls, the CPC (calling party category) code in the IAM (Section 9.2.1) is 001101. After the test, exchange W can unblock the trunk by sending an UBL (Unblocking) signal to exchange V, which then sends an UBA (Unblocking Acknowledgment) signal and resumes seizing trunk T for new calls. TUP trunks are usually bothway trunks, and both exchanges can therefore initiate the blocking of a trunk.

9.5.3 Resetting a Trunk

Memory devices of the processors in stored-program controlled exchanges store records with status information for each attached trunk. When a memory malfunction causes exchange W to lose the status information of a trunk T of group TG, it sends a RSC (Reset Circuit) signal—Table 9.5-1—for the trunk to exchange V. This signal requests V to clear a possible connection involving T and set the trunk to "idle."

If the memory at V indicates that T is currently involved on an incoming call (from W), the exchange regards the RSC as a clear-forward signal. It releases trunk T, sends a CLF (Clear-Forward) signal to the next exchange in the connection, and sends a RLG (Release-Guard) signal for trunk T to exchange W (Table 9.2-1).

TABLE 9.5-1 Circuit Supervision Signals (H0 = 0111)

H1	Name	Acronym
0001	Release-Guard	RLG
0010	Blocking	BLO
0011	BLO Acknowledgment	BLA
0100	Unblocking	UBL
0101	UBL Acknowledgment	UBA
0110	Continuity Check Request	CCR
0111	Reset Circuit	RSC

Source: Rec. Q.730. Courtesy of ITU-T.

If the record at V shows that its call on T is an outgoing call (to W), it sets trunk T to "idle," sends a CLB signal to the previous exchange in the connection, and sends a CLF signal to W, which then responds with a RLG signal. If exchange V receives a RSC signal when the affected trunk T is idle, it responds with a RLG signal.

9.5.4 Circuit Group Blocking, Unblocking, and Resetting

TUP signaling includes a group of *circuit group supervision messages* that are used to block, unblock, and reset up to 256 trunks in a trunk group simultaneously. The message acronyms and heading codes are listed in Table 9.5-2.

The structure of a circuit group supervision message is shown in Fig. 9.5-2. The values of OPC and DPC (originating and destination point codes) identify the trunk group TG. The particular set of up to 256 trunks that are covered by the message is identified by the values of parameters CIC (circuit identification code) and R (range). For example, the combination (CIC $= c$, R $= r$) identifies the set of trunks in group TG that have the CIC values c, $c+1$, $c+2, \ldots$, and $c+r$.

In Fig. 9.5-2, parameter R is followed by up to 32 octets of *status indicator* bits for the trunks in the set. Bits 1 through 8 in octet 3 represent the trunks with CIC values $(c+0)$ through $(c+7)$, bits 1 through 8 in the next octet represent trunks with CIC codes $(c+8)$ through $(c+15)$, and so on.

In the request messages, the indicator value is 1 if the request is made for the particular trunk, and 0 otherwise. These indicator values are copied in the group blocking and unblocking acknowledgment messages.

In the Group Reset Acknowledgment (GRA) message, a status indicator value 1 or 0 indicates that the corresponding trunk is blocked, or not blocked, by the exchange that sends the GRA message.

Maintenance and Failure Group Blocking. A Maintenance Group Blocking (MGB) message is sent by the exchange (V or W) that wants to test a number of trunks in trunk group TG. A Hardware Failure Group Blocking (HGB) message is sent by an exchange when it detects a failure on a number of trunks.

TABLE 9.5-2 Circuit Group Supervision Messages (H0 = 1000)

H1	Name	Acronym
0001	Maintenance Group Blocking Request	MGB
0010	Maintenance Group Blocking Acknowledgment	MGA
0011	Maintenance Group Unblocking Request	MGU
0100	Maintenance Group Unblocking Acknowledgment	MUA
0101	Hardware Failure Group Blocking Request	HGB
0110	Hardware Failure Group Blocking Acknowledgment	HBA
0111	Hardware Failure Group Unblocking Request	HGU
1000	HGU Acknowledgment	HUA
1001	Group Reset Request	GRS
1010	Group Reset Acknowledgment	GRA

Source: Rec. Q.723. Courtesy of ITU-T.

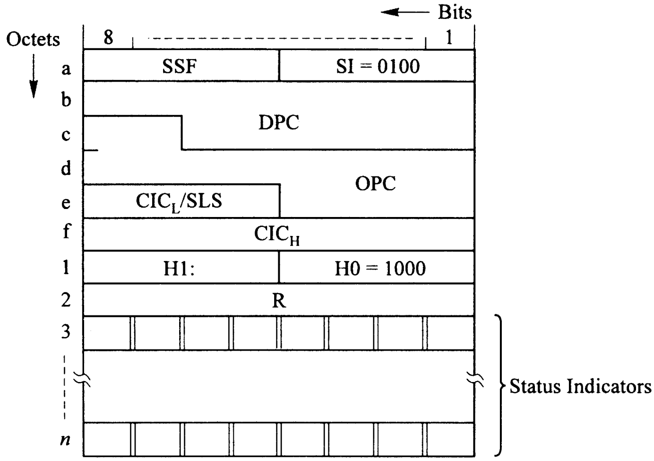


Figure 9.5-2. MTP3 message with circuit group supervision message. (From Rec. Q.723. Courtesy of ITU-T.)

When an exchange receives a Maintenance (MGB) or Hardware Failure (HGB) Group Blocking message, it stops seizing the affected trunks for new calls. Existing calls are allowed to continue after receipt of a MGB but are terminated immediately when a HGB is received.

A group reset or group blocking request can affect calls on a large number of trunks. Therefore, as a safety measure, the requesting exchange sends the message twice, and the other exchange takes action only if it receives two identical MGB or HGB messages in a 5-s interval.

9.5.5 Signaling Route Set Congested or Not Available

The TUP process at an exchange responds to MTP-pause, MTP-resume, and MTP-status indications from its MTP (Sections 8.9.3 and 8.9.5). We again consider the trunk group TG between exchanges V and W and denote the TUP process at exchange V by TUP-V.

When TUP-V receives a MTP-pause indication for destination W, it knows that it can no longer send and receive signaling messages for the trunks in group TG. It therefore clears all existing calls in TG and suspends the seizure of trunks for new calls. When TUP-V receives a MTP-resume for destination W, it resumes seizing trunks in group TG.

When TUP-V receives a MTP-status indication for destination W, it reduces the number of seizures of trunks in group TG. The reduction remains in effect until a period of about 5 s has elapsed during which TUP-V has not received a new status indication for destination W.

9.6 VERSIONS OF TUP SIGNALING

ITU-T Recommendations Q.721–Q.724 [1–4] are useful as a general framework for TUP signaling systems. However, many telecoms have found it necessary to tie down the "loose ends," select options, and make adaptations to accommodate the services and procedures that already existed in their respective national networks prior to the introduction of TUP. As a result, there are differences in national TUP versions. This situation is comparable to the national versions of national R2 signaling.

Moreover, in the 1984–1988 time frame, several telecoms began to introduce some ISDN services, which had not yet been standardized by ITU-T (ISDN services and ISUP interexchange signaling only started to be adequately defined in the 1989 CCITT *Blue Books*). To handle these "pseudo-ISDN" services, some telecoms have defined and deployed "extended" versions of TUP.

This section briefly describes a number of TUP versions.

9.6.1 British Telecom National User Part

British Telecom has defined a national version of SS7 signaling originally called *CCITT No. 7 (BT)*, for use in the United Kingdom. It consists of a slightly modified version of the ITU-T-defined Message Transfer Part and a National User Part (NUP), which supports, in addition to basic telephone call control, some supplementary services and some initial ISDN features [7].

NUP transcends TUP in that it handles calls between analog subscriber lines and calls between subscribers who have digital access lines that use a signaling system called Digital Access Signaling System (DASS), of which there are two versions.

NUP message coding resembles ITU-T TUP coding. However, headings H0 and H1 occupy one octet each, thus leaving room for many additional (future) messages.

Sending of Called Number. NUP includes three address messages. Initial address messages carry the leading digits of the called address only (for overlap address signaling). After receipt of an IAM, an exchange Y can request a specific number (*N*) of subsequent digits from preceding exchange X. This is an established signaling procedure in the United Kingdom. The request is made with a Send-N-Digits message. As soon as X has received the requested digits, it sends them in a subsequent address message (SAM). The last subsequent address message is called a Final Address Message (FAM). The Initial and Final Address Message (IFAM) carries a complete called number and is used for *en-bloc* signaling.

The initial (IAM or IFAM) message of a call includes a number of other parameters, some of which are not available in ITU-T-defined TUP. Notably, *signaling handling protocol* (SHP) indicates whether the call originated from an analog or digital subscriber line, and *call path indicator* (CPI) indicates whether an analog or 64-kb/s digital path is required for the call.

End-to-End Signaling. During the call setup, the calling and called local exchanges can communicate to obtain additional information. A number of *service information messages* (SIMs) have been defined for this purpose. Their functions are similar to those of the GRQ and GSM messages described in Section 9.4.2.

Swap Message. Two digital subscribers can change their mode of communication from speech to digital data and vice versa during a call. The subscriber originating the change signals his local exchange X with a DASS *Swap* message. This information is transferred across the network to the distant local exchange Y by NUP *Swap* messages and delivered to the other subscriber as a DASS message.

9.6.2 France Telecom TUP-E

France Telecom recognizes three types of subscribers: conventional telephone subscribers, Transcom-ISDN subscribers, and ISDN subscribers [8]. Calls from telephone subscribers are handled on trunks with Socotel MF, basic (French) TUP, or Extended *TUP* (TUP-E) signaling. Transcom-ISDN subscribers can make 64-kb/s calls (only). These calls are set up by "telephone" subscriber signaling and have to be routed on digital trunks with TUP-E signaling. ISDN subscribers use a national form of ISDN access signaling and can make speech and data calls, which are also routed on digital trunks with TUP-E signaling exclusively.

The purpose of TUP-E is similar that of British Telecom's NUP signaling. It has allowed the introduction of a number of ISDN services at a time that ITU-T had not yet defined digital subscriber signaling (DSS) and ISUP (ISDN interexchange signaling) in sufficient detail.

9.6.3 TUP in the International Network

Basic International TUP. TUP signaling for basic call control, as described in Section 9.3, is used on a number of international (mostly transatlantic and transpacific) trunk groups [9]. This version of TUP also supports international digital connectivity (transparent transfer of 64-kb/s user communications)—see Section 9.4.4.

TUP+. TUP+ has been introduced on international circuits between countries of the European Economic Community. It was defined by the CEPT (now ETSI) in its Recommendation 43-02 [10]. It includes signals and procedures to support a number of ISDN and supplementary services, such as closed user groups, calling line identification, and user-to-user signaling.

9.7 ACRONYMS

ACB	Access Barred signal
ACC	Automatic Congestion Control message

ACM	Address Complete message
ADI	Address Incomplete signal
ANC	Answer, Charge signal
ANN	Answer No Charge signal
ANSI	American National Standards Institute
ANU	Answer (no instructions on charging) signal
BLA	Blocking Acknowledgment signal
BLO	Blocking signal
CAS	Channel-associated signaling
CBK	Clear-Back signal
CCITT	International Telegraph and Telephone Consultative Committee
CCF	Continuity Failure signal
CCIS	Common-Channel Interoffice Signaling
CCR	Continuity Check Request signal
CEPT	European Conference of Postal and Telephone Administrations
CFL	Call Failure signal
CGC	Circuit Group Congestion signal
CIC	Circuit Identification Code
CIC_L	Four low-order bits of CIC
CIC_H	Eight high-order bits of CIC
CLF	Clear-Forward signal
COT	Continuity signal
CPC	Calling party category
CPI	Call path indicator
CUG	Closed user group
DASS	Digital Access Signaling System
DPC	Destination point code
DPN	Digital Path not Provided signal
ETSI	European Telecommunications Standards Institute
FAM	Final Address message
FDM	Frequency-division multiplexing
FOT	Forward-Transfer signal
GRA	Group Reset Acknowledgment message
GRQ	General Request message
GRS	Group Reset Request message
GSM	General Forward Setup Information message
H0	Message heading field 0
H1	Message heading field 1

HBA	Hardware Failure Group Blocking Acknowledgment message
HGB	Hardware Failure Group Blocking Request message
HGU	Hardware Failure Group Unblocking Request message
HUA	Hardware Failure Group Unblocking Acknowledgment message
IA	Incoming access
IAI	Initial Address Message with Additional Information
IAM	Initial Address message
ICB	Incoming calls barred
IFAM	Initial and Final Address message
ISC	International switching center
ISDN	Integrated services digital network
ISUP	ISDN user part of SS7
ITU	International Telecommunication Union
ITU-T	Telecommunication Standardization Sector of ITU
LOS	Line Out of Service
MFC	Multifrequency signaling Compelled
MGA	Maintenance Group Blocking Acknowledgment message
MGB	Maintenance Group Blocking Request message
MGU	Maintenance Group Unblocking Request message
MSU	Message signal unit on signaling link
MTP	Message Transfer Part
MTP3	Message transfer part level 3
MUA	Maintenance Group Unblocking Acknowledgment message
NICB	No incoming calls barred
ND	Number of digits
NIA	No incoming access
NNC	National Network Congestion signal
NOA	No outgoing access
NUP	National User Part
OA	Outgoing access
OPC	Originating Point Code
PAD	Point code of the affected destination
PCM	Pulse code modulation
RAN	Re-Answer signal
RL	Routing label
RLG	Release-Guard signal
RSC	Reset Circuit signal
SAM	Subsequent Address message

SAO	Subsequent One-Digit Address Message
SEC	Switching Equipment Congestion signal
SHP	Signaling Handling Protocol
SI	Service indicator
SIF	Signaling information field
SIM	Service information message
SIO	Service information octet
SL	Signaling link
SLS	Signaling link selector
SSB	Subscriber Busy
SSF	Subservice field
SST	Send Special Information Tone signal
ST	End of address signaling
SLS	Signaling link selector
SS6	Signaling system No.6
SS7	Signaling system No.7
TUP	Telephone User Part
UBA	Unblocking Acknowledgment signal
UBL	Unblocking signal
UNN	Unallocated Number signal

9.8 REFERENCES

1. *Specifications of Signalling System No. 7*, Rec. Q.721, ITU-T, Geneva, 1988.
2. *Specifications of Signalling System No. 7*, Rec. Q.722, ITU-T, Geneva, 1988.
3. *Specifications of Signalling System No. 7*, Rec. Q.723, ITU-T, Geneva, 1988.
4. *Specifications of Signalling System No. 7*, Rec. Q.724, ITU-T, Geneva, 1988.
5. *Specifications of Signalling System No. 7*, Rec. Q.730, ITU-T, Geneva, 1988.
6. D. R. Manfield, G. Millsteed, and M. Zukerman, "Congestion Controls in SS7 Signaling Networks," *IEEE Commun. Mag.*, **31**, No. 6, June 1993.
7. CCITT Signalling System No. 7: National User Part, *Br. Telecommun. Eng.*, **7**, Apr. 1988.
8. J. Craveur et al., CCITT No. 7 Common Channel Signalling in the French Telecommunication Network, *ISS Conf. Rec.*, Phoenix, 1987.
9. J. J. Lawser et al., Common Channel Signalling for International Services, *ISS Conf. Rec.*, Stockholm, 1990.
10. T/S 43-02E, CEPT, 1988 (now ETSI ETR 256).

10

DIGITAL SUBSCRIBER SIGNALING SYSTEM NO. 1

Digital Subscriber Signaling System No. 1 (DSS1) is used for signaling between an ISDN (integrated services digital network) subscriber and the local exchange.

This chapter is a brief excursion from the description of Signaling System No. 7 (SS7), which has been the subject of Chapters 7 through 9, and will continue in Chapter 11, with the description of the Integrated Services Digital Network User Part (ISUP). This order of presentation has been chosen because DSS1 can be discussed without references to ISUP, and because it is helpful to be familiar with DSS1 when exploring ISUP, which is the interexchange signaling system for the control ISDN interexchange calls.

10.1 INTRODUCTION TO ISDN AND DSS1

This section presents a broad-brush introduction to ISDN, focusing primarily on its signaling aspects. More details on ISDN architecture and technology can be found in a number of texts [1–4].

10.1.1 Circuit-Mode and Packet-Mode Networks

So far, we have discussed (circuit-mode) telecommunication networks in which two subscribers communicate over a temporary and dedicated circuit, consisting of the subscriber lines and possibly one or more trunks. Connections are set up by exchanges at the start of a call and released when the call has ended.

During the 1970s, a new network architecture was developed for data communications. These networks, which have been installed in a number of countries, consist of *nodes* that are interconnected by *data links*. The users of these network have *data*

Signaling in Telecommunication Networks, Second Edition, by John G. van Bosse and Fabrizio U. Devetak

terminal equipment (DTE) at their premises, which are attached to one of the nodes. Users communicate by sending *packets*, consisting of a *header* and a limited amount of data (say, up to 100 octets). The packets are transferred on the data links, and each node uses the address information in the packet header to select an outgoing data link to or toward the packet destination. A data link is not "dedicated" to the communications between a specific pair of users: it transfers packets of several users.

This type of communication is known as *packet-switched* (or *packet-mode*) communication. ITU-T has defined standards for packet data networks in recommendations of the "X" series. In particular, Rec. X.25 [5] defines the interface between the DTE and the data communication network.

We have already encountered an example of such a network. The signaling networks described in Chapter 5 are packet-mode networks for a particular application: the transfer of common-channel signaling messages between exchanges (and other entities in the telecommunication network). Even though the terminology is different, the concepts are the same:

Concept	**Data Communication Network**	**Signaling Network**
Node	Node	Signal transfer point
Link	Data link	Signaling data link
Data unit	Packet	Signal unit
User	Customer's DTE	Signaling end point

Circuit-mode and packet-mode data communications exist side by side. Packet-mode is the preferred choice for data traffic that consists of short bursts that are separated by comparatively long "silent" intervals. This is often the case for communications between data terminals and a centralized mainframe computer. For example, the verification of credit cards by stores and the reservation of hotel rooms by travel agencies.

Packet communication is discussed in Chapter 20.

10.1.2 The Objectives of ISDN

When discussing ISDN, it is customary to speak about *users* instead of subscribers. ISDN has two objectives. In the first place, it is an "integrated services network" that provides circuit-mode (speech and data) and packet-mode (data) communications for its users.

The second objective is *user-to-user digital connectivity*. This means that all components (lines, trunks, exchanges) in an ISDN connection transfer 64-kb/s digital data.

10.1.3 User Equipment and Access

Figure 10.1-1 shows the equipment at the premises of an ISDN user and the access to the ISDN network. The user's *network termination* (NT) is connected to a number of

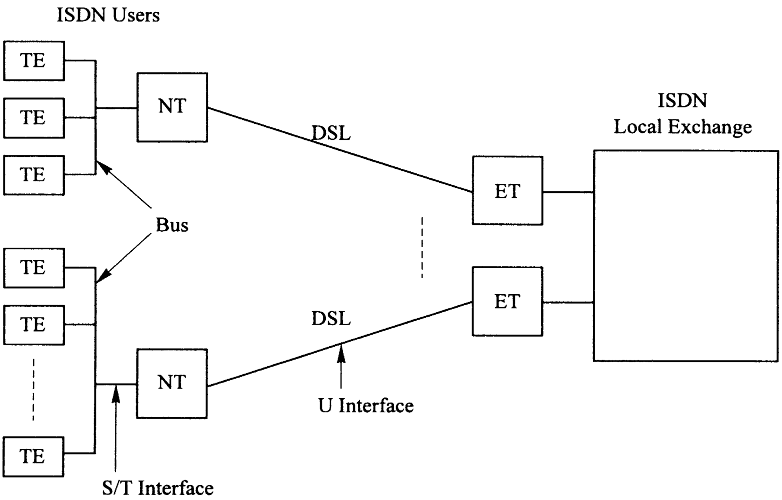


Figure 10.1-1. ISDN users served by a local ISDN exchange. TE, Terminal equipment; NT, network terminal; DSL, digital subscriber line; ET, exchange terminal.

terminal equipment (TE) units of various types (telephones, facsimile machines, data processing equipment, etc.). The NT is attached to an *exchange termination* (ET) in the local exchange by a *digital subscriber line* (DSL). The line has a number of time-division multiplexed B-channels and one D-channel. The B-channels are used for circuit-mode communications (PCM encoded speech or voice-band modem data, or 64-kb/s digital data). The D-channel is used for DSS1 signaling messages and for packet-mode data communications.

In most countries, the DSL, NT, and ET are considered to be internal parts of the telecommunication network. Therefore, the network-to-user interface specified in ITU-T is the S/T interface (between NT and the TEs). In the United States, the DSL and NT are considered to be outside the network, and the network-to-user interface is the U interface. We distinguish the following two DSL types.

Basic Access. A basic access DSL is intended for residential or small business use. In the United States, it is a two-wire circuit with two bidirectional B-channels (64 kb/s) and one bidirectional D-channel that operates at 16 kb/s (2B + D). The bit rate in each direction is $2 \times 64 + 16 + 48$ (for overhead functions) $= 192$ kb/s. Physically, the DSL is an ordinary two-wire subscriber line. Circuitry in the ET and NT enables the line to transfer both bit streams simultaneously (full duplex operation). Since a call requires one B-channel, this DSL allows two simultaneous calls, and up to eight TEs can be connected to the NT.

Other countries use a variety of basic access DSLs. They can be two- or four-wire circuits with one (B + D) or two (2B + D) B-channels.

At the user's premises, the TEs are attached to the NT by a bidirectional *passive bus*. Each terminal has access to the B-channels and the D-channel of the DSL.

Primary Rate Access. Primary rate access DSLs are intended for medium or large businesses and resemble the first-order digital multiplex transmission systems that have been developed for PCM trunks (Section 1.5.2). These DSLs consist of two amplified two-wire channels, one for each direction of transmission.

American primary rate DSLs have the channel format of the T1 digital transmission system. They operate at 1544 kb/s and have twenty-four 64-kb/s channels. Channels 1–23 are used as B-channels, and channel 24 is the (64-kb/s) D-channel (23B + D). Most other countries use a digital transmission system that operates at 2048 kb/s. Channels 1–15 and 17–31 are used as B-channels and channel 16 is the D-channel (30B + D).

10.1.4 The ISDN Network

From the user's point of view, ISDN is an "integrated services" network. The user's DSL can be used for circuit- and packet-mode communications. Before the arrival of ISDN, a subscriber needed a line to his local exchange for circuit-mode communications, and a separate line to a node in a data communications network for packet-mode communications.

Actually, only the DSLs and the local ISDN exchanges are truly integrated. As shown in Fig. 10.1-2, an ISDN network is a group of the following three networks.

Telecommunication (Circuit-Switched) Network. A circuit-switched "pure" ISDN network would consist of 64-kb/s digital trunks and exchanges with digital switchblocks only. In practice, these networks are gradual evolutions of existing telecommunication networks. They include a mixture of analog (FDM) and digital trunks, and exchanges with analog or digital switchblocks. The analog equipment is gradually being replaced by digital equipment. The networks usually have interexchange signaling systems of several types but have mostly been converted to ISUP signaling. This system, described in Chapter 11, is the only system that meets the requirements of ISDN.

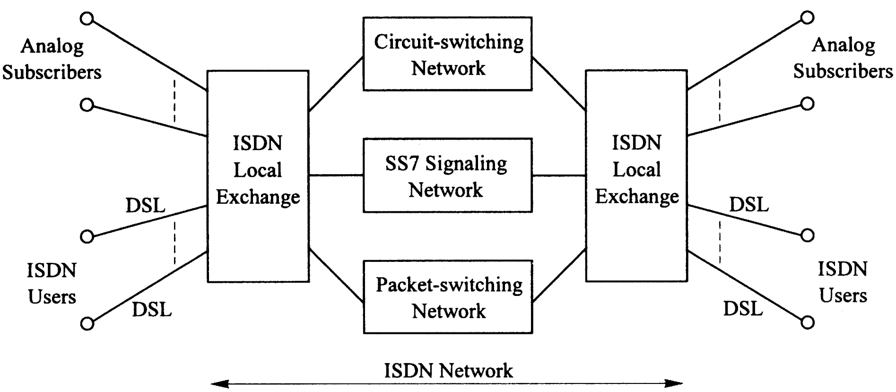


Figure 10.1-2. ISDN network.

Since the number of ISDN users is very small compared to the number of (analog) subscribers, only a small fraction of the local exchanges in a network is equipped to serve ISDN users, and these exchanges also serve analog subscribers.

SS7 Signaling Network. This may be an extension of an existing SS7 signaling network (if the telecommunication network is already using TUP signaling) or a newly installed network.

Packet-Switching Network. This may be an existing data communication network, adapted with an interface to ISDN local exchanges, or a new network.

The local exchange segregates the DSS1 signaling messages and data packets that arrive on the D-channels of a DSL. The packets are transferred to the packet-switching network, and the incoming DSS1 signaling messages are processed by the exchange. In the reverse direction, the local exchange transfers the packets received from the packet-switching network, and its outgoing DSS1 signaling messages, to the D-channels of the DSLs.

10.1.5 ISDN Services for Circuit-Mode Communications

This book covers signaling in circuit-switched networks. Therefore, this chapter describes the circuit-mode communications of ISDN only. Three groups of ISDN services for this mode of communications are listed below.

Bearer Service. This defines the type of communication service for a call. A calling ISDN user can request one of three bearer services: speech, 3.1-kHz audio (voiceband modem data), or 64-kb/s digital data. Information about the type of bearer service is transferred by the network to the called user, where it is one of the criteria to select an appropriate TE. For example, when the bearer service is "speech," the incoming call is connected to a telephone. We shall see in Chapter 11 that the bearer service is also taken into account by the exchanges in the network for the selection of outgoing trunks. For example, speech calls can be set up on analog or digital trunks, but 64-kb/s data calls require digital trunks.

Teleservice. The data in 3.1-kHz audio and 64-kb/s calls can pertain to various data services (facsimile, telex, teletext, etc.). The calling user specifies the type of teleservice for the call. The teleservice information is transferred transparently (i.e., without examination) by the network. It is processed by the called user equipment, to select the appropriate TE for the incoming call.

Supplementary Services. These services vary from country to country. In countries that use TUP signaling, the supplementary services supported by TUP (malicious call handling, calling line identification, call forwarding, and closed user group service) (Section 9.4)—that are available to (analog) subscribers—are also available to ISDN users.

In addition, ISDN can provide *user-to-user signaling*. This allows two ISDN users to send signaling messages to each other during the setup and clearing phases of a call. The network transfers this information transparently.

Closed user group service and user-to-user signaling are not offered in the United States. On the other hand, the ISDN for the United States includes a number of services for multilocation businesses.

10.1.6 Introduction to DSS1

DSS1 is a message-oriented signaling system [1–5]. Literature on DSSl is usually in terms of signaling between a *user* and the *network*. Actually, the signaling takes place between a TE of an ISDN user and the local exchange to which the user's DSL is attached. The DSS1 signaling messages are carried in the D-channel of the DSL, which is the common signaling channel for the TEs on a DSL.

DSS1 and Signaling System No. 7 (SS7) have been specified by different ITU-T study groups and use different "languages" (terms). However, many DSS1 concepts are similar to SS7 concepts.

Figure 10.1-3 shows a local exchange, a SS7 signaling link, and a DSL. The functions of the D-channel are comparable to those of the SS7 signaling link. The information units on the D-channel, which are known as *frames*, are similar to the signal units (SUs) of SS7.

SS7 is organized as a hierarchy of protocols. The message transfer part (MTP) serves a number of SS7 user parts, such as the telephone user part (Chapter 9) and ISUP (Chapter 11). In a similar fashion, DSSI is divided into the *data link*

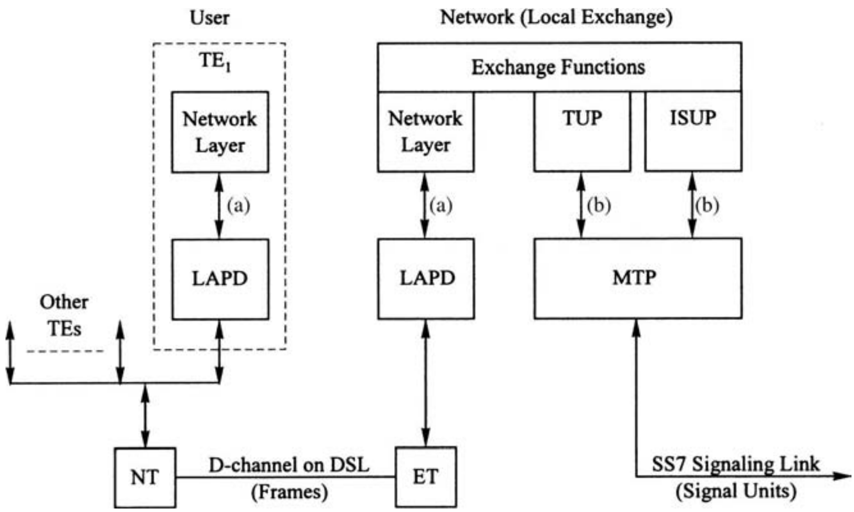


Figure 10.1-3. Functional entities in DSS1 and ISUP signaling. (a), DSS1 primitives; (b), SS7 primitives.

layer, also known as LAPD (Link Access Protocol for D-channel), and the *network layer*.

The functions of LAPD are comparable to those of MTP. The network layer includes protocols comparable to those of ISUP. The network-layer protocols are usually referred to as "Q.931 protocols" because they have been specified in ITU-T Recommendation Q.931.

As in SS7, LAPD and the network layer communicate by passing "primitives" (Section 7.3).

10.2 DATA LINK LAYER (LAPD)

The primary function of LAPD is the reliable transfer of frames between a TE and the local exchange [1–7]. It includes provisions for error detection and correction.

10.2.1 Data Link Connections

We start by describing the functional entities at the ISDN user and at the exchange in more detail. Figure 10.2-1 shows a user with two terminals at his premises.

Each TE on a DSL is identified by a *terminal endpoint identifier* (TEI), which has a value in the range 0–126. The TEs in this example have TEI values 1 and 2.

A terminal has two LAPD functions. One is TE-specific, and identified by the TEI of the terminal. The other one is identified by TEI = 127 at all terminals. Each terminal LAPD has a "peer" LAPD at the exchange.

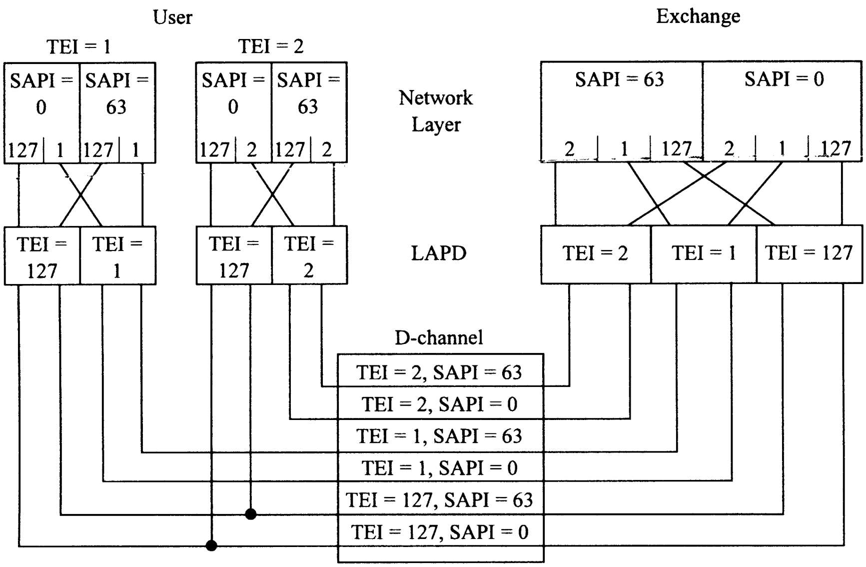


Figure 10.2-1. Data link connections on a D-channel.

A TE has several network-layer functions. Each function is identified by a *service access point identifier* (SAPI). This chapter considers the following functions only:

SAPI	Function
0	ISDN call control
63	Management of data link layer

These functions are present in each TE and have peers in the exchange. Frames are always transferred between a terminal LAPD and the peer LAPD at the exchange. Moreover, a frame that originates at a network-layer function at one end of the D-channel is delivered to the peer function at the other end.

There are a number of *data link connections* on a D-channel. Each connection is identified by a combination of TEI and SAPI values. The connections with TEI = 0–126 are bidirectional point-to-point connections, between a TE on a DSL and its "peer" function at the exchange. For example, the connection (TEI = 2, SAPI = 0) carries call-control frames to and from the terminal identified by TEI = 2.

The connections with TEI = 127 are "point-to-multipoint" in the direction from exchange to the user. All TEs on a DSL examine received frames with this TEI value. An exchange can broadcast a message to all TEs on a DSL, by sending a frame with TEI = 127.

10.2.2 Frames

The general format of LAPD frames is shown in Fig. 10.2-2. Frames are separated by one-octet flags. The flag pattern (0111 1110) is the same as in SS7.

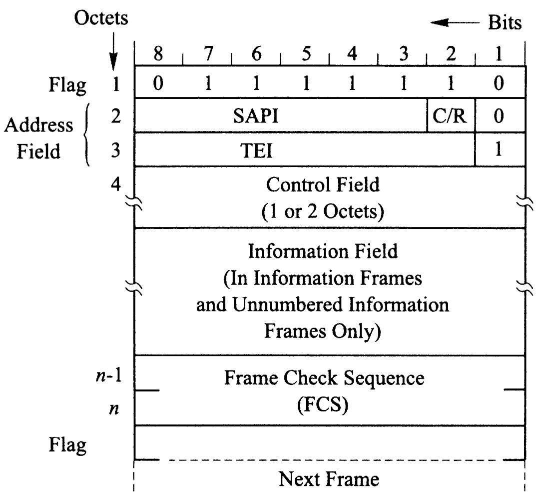


Figure 10.2-2. General LAPD frame format. (From Rec. Q.921. Courtesy of ITU-T.)

The *address field* (octets 2 and 3) of a frame contains SAPI and TEI and is used to route the frame to its destination. The *control field* starts in octet 4 and consists of one or two octets. The *information field* is present in some frame types only.

Octets $n - 1$ and n contain a 16-bit *frame check sequence* (FCS) field. It has the same function as the CB (check-bit) field in SS7 signal units (Section 8.3.2) and enables a LAPD to detect errors in a received frame.

10.2.3 Frame Types

We distinguish the following frame types—see Table 10.2-1.

Information Frame (I). This frame is comparable with the message signal unit of SS7. It has an information field that carries a network-layer message to/from a particular TE on a DSL. The TE is identified by the value of TEI (0 through 126).

Supervision Frames (Group S). These frames originate at the LAPD at one end of a data link connection, are processed by the LAPD at the other end, and contain information on the status of the connection. These frames do not have an information field and are comparable with the link status signal units of SS7 (7.3.1).

Unnumbered Frames (Group U). These frames have no counterparts in SS7. The Unnumbered Information (UI) frame is the only frame in this group that has an information field and carries a network-layer message. To broadcast a message to all TE on a DSL, the exchange sends an UI frame with TEI = 127.

10.2.4 Control Fields, C/R, P, and F Bits

The control fields of the various frames consist of one or two octets (Table 10.2-1). Bits 1 and 2 of octet 4 identify a group of frames:

Frame	Octet 4, Bit 2	Octet 4, Bit 1
Information frame	0 or 1	0
Group S (supervision frames)	0	1
Group U (unnumbered frames)	1	1

In supervision and unnumbered frames, other bits in octet 4 identify a particular frame type within the group.

The C/R bit in the address field and the P or F bit in the control field have been carried over from the X.25 protocol, which has been specified by ITU-T for data communication networks [5]. These bits are set by the LAPD at one end of a connection and are processed by its peer.

TABLE 10.2-1 Control Fields

Group[a]	Frame Type	C	R	8	7	6	5	4	3	2	1	Octet
				Control Field (bits)								
	I Information	X		N(S)							0	4
				N(R)							P	5
S	RR Receive Ready	X	X	0	0	0	0	0	0	0	1	4
				N(R)							P/F	5
S	RNR Receive Not Ready	X	X	0	0	0	0	0	1	0	1	4
				N(R)							P/F	5
S	REJ Reject	X	X	0	0	0	0	1	0	0	1	4
				N(R)							P/F	5
U	UI Unnumbered Information	X		0	0	0	P	0	0	1	1	4
U	SABME Set Automatic Balanced Mode Extended	X		0	1	1	P	1	1	1	1	4
U	DM Disconnected Mode		X	0	0	0	F	1	1	1	1	4
U	DISC Disconnect	X		0	1	0	P	0	0	1	1	4
U	UA Unnumbered Acknowledgment		X	0	1	1	F	0	0	1	1	4

[a]S, Supervision frames; U, unnumbered frames.

Source: Rec. Q.921. Courtesy of ITU-T.

The value of C/R (Fig. 10.2-2) classifies each frame as a *command* or a *response* frame:

Frame	Frames Sent by Exchange	Frames Sent by Terminal
Command frame	C/R = 1	C/R = 0
Response frame	C/R = 0	C/R = 1

The I, UI, SABME, and DISC frames are command frames, and the DM and UA frames are response frames. The supervision frames can be sent as command or response frames.

By setting the P bit in a command frame to P = 1, a LAPD orders its peer to respond with a supervision or unnumbered frame. A response frame with F = 1 indicates that it is sent in response to a received command frame in which P = 1.

10.2.5 Transfer Modes

LAPD transfers frames in one of the following two modes.

Multiframe Acknowledged Message Transfer. This mode is used only on point-to-point data link connections, for the transfer of I frames. It includes error correction by retransmission and in-sequence delivery of error-free messages. This mode is similar to the basic error correction of *message signal units* (MSUs) in SS7 (Section 8.4).

The control field of an I frame has a *send-number* [N(S)] field and a *receive-number* [N(R)] field—see Table 10.2-1. These fields are comparable to the forward and backward sequence number fields (FSN, BSN) in MSUs. A LAPD assigns increasing "send" sequence numbers N(S) to consecutive transmitted I frames: $N(S) = 0, 1, 2, \ldots, 127, 0, 1, \ldots,$ etc.). It also stores the transmitted frames in a retransmission buffer where they are kept and are available for retransmission, until positively acknowledged by the distant LAPD.

We examine the transfer of I frames on a data link connection, sent from terminal to exchange—see Fig. 10.2-3. The procedure in the other direction is the same. LAPD-E and LAPD-T denote the data link functions at the exchange and terminal, respectively.

LAPD-E checks all received frames for errors. In addition, error-free I frames are "sequence checked." If the value of the N(S) is one higher (modulo 128) than the N(S) of the most recently accepted I frame, the new frame is "in-sequence" and therefore accepted, and its information field is passed to the specified network-layer function.

LAPD-E acknowledges accepted I frames with the receive number [N(R)] in its outgoing I and supervision frames. The value of N(R) is one higher (modulo 128) than the value of N(S) in the latest accepted I frame.

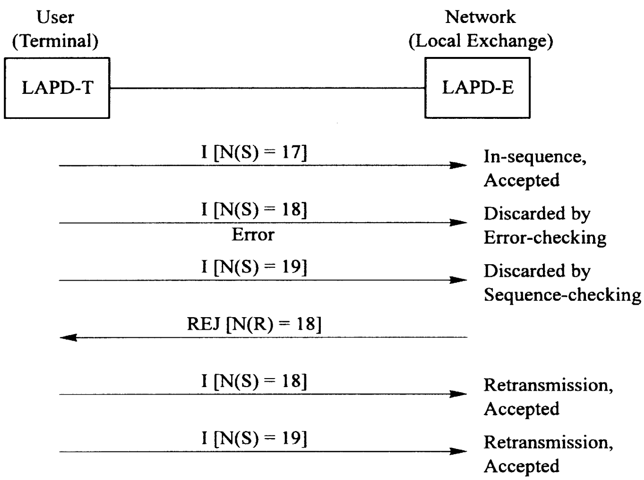


Figure 10.2-3. Error correction on information frame.

Suppose that the most recently accepted I frame had N(S) = 17, and that the I frame with N(S) = 18 incurred a transmission error and was discarded by LAPD-E. The next I frame [with N(S) = 19] passes LAPD-E error-checking but arrives out-of-sequence and is discarded by LAPD-E sequence-checking. LAPD-E now sends a Reject (REJ) frame with N(R) = 18. This requests LAPD-T to retransmit the I frames in its retransmission buffer, starting with the frame with N(S) = 18. At LAPD-E, the sequence-checking continues to discard I frames until it receives the (retransmitted) frame with N(S) = 18.

The two message streams (terminal to network, and vice versa) in a point-to-point data link connection are independent of each other and independent of the message streams in the other point-to-point connections in the D-channel. On a D-channel with n point-to-point connections, there are $2n$ independent N(S)/N(R) sequences.

Unacknowledged Message Transfer. Supervision (S) and unnumbered (U) frames do not include a N(S) field. They are accepted when received without errors and are not acknowledged. Supervision frames include a N(R) field to acknowledge received information frames.

Unnumbered Information (UI) frames do not include N(S) and N(R) fields, because they are always sent with the "group" TEI (127), and it is not possible to coordinate send and receive sequence numbers for the "group" functions in the terminals on a DSL.

Primitives. A network-layer function requests the acknowledged mode for an outgoing message by passing it in a *DL-data* request primitive to its LAPD, which then transfers the message in an I frame. Unacknowledged transfer is requested by passing the message in a *DL-unitdata* request, and the message is then carried in a UI frame.

10.2.6 Supervision of Data Link Connections

The LAPDs at the ends of a point-to-point data link connection can send and receive the following supervision frames:

RR (*Receive Ready*) is sent by a LAPD to indicate that it is ready to receive I frames.

RNR (*Receive Not Ready*) is sent by LAPD to indicate that it is not able to receive I frames but will process received supervision frames.

REJ (*Reject*) indicates that the sending LAPD has rejected a received I frame.

A supervision action can originate at either end of the connection. Figure 10.2-4 shows a few examples and illustrates the use of the C/R, P, and F bits. LAPD-T and LAPD-E denote the data link layer functions at the terminal and the network side of the connection.

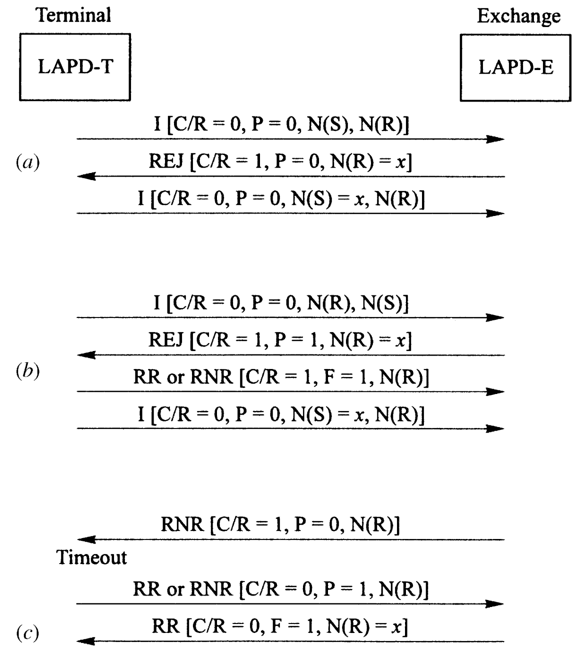


Figure 10.2-4. Examples of supervision procedures.

In example (a), LAPD-E has received an out-of-sequence I frame and rejects it with a REJ command frame in which P is set to 0 (no acknowledgment required). $N(R) = x$ indicates that the last accepted I frame had $N(S) = x - 1$. LAPD-T then retransmits the I frames in its retransmission buffer, starting with the frame whose N(S) equals x.

Example (b) deals with the same situation, except that LAPD-E has set P to 1 in its REJ command frame. This orders LAPD-T to acknowledge the frame. LAPD-T therefore first sends a RR or RNR response frame ($C/R = 1$, $F = 1$) and then starts the retransmission of I frames.

In example (c), LAPD-E indicates that it is unable to receive I frames, with a RNR command frame. LAPD-T suspends sending I frames and starts a timer. If it receives a RR frame before the timer expires, it resumes transmitting or retransmitting I frames.

If the timer expires and no RR frame has been received, LAPD-T sends a "command" supervision frame ($C/R = 1$), with P set to 1. This orders LAPD-E to send a supervision "command" frame. In the example, LAPD-E responds with a RR frame, indicating it is ready again to accept I frames, and that the last accepted I frame had $N(S) = x - 1$. LAPD-T then resumes transmitting or retransmitting I frames, starting with the frame with $N(S) = x$.

If the response of LAPD-E had been a RNR frame, LAPD-T would have restarted its timer and waited again for a RR frame. If LAPD-E then remains receive-not-ready after several timeouts, LAPD-T informs its network-layer function.

10.2.7 SABME, DISC, DM, and UA Frames

These unnumbered frames (Table 10.2-1) are used to start and end multiframe acknowledged operation on a point-to-point data link connection.

For example, when the network-layer function identified by SAPI = 0 at an exchange needs to initiate multiframe acknowledged operation on a DSL connection identified by, say, TEI = 5, SAPI = 0, it passes a Set Automatic Balanced Mode Extended (SABME) frame to the LAPD-E for the connection, which then initializes its N(S) and N(R) counters to 0 and transfers the frame to the peer LAPD-T. That LAPD then initializes its counters, informs the SAPI = 0 network-layer function, and returns an Unnumbered Acknowledgment (UA) frame to the exchange. The LAPD-E then informs the requesting network-layer function that it can start sending I frames.

An existing multiframe acknowledgment operation on a connection is terminated by the sequence of a Disconnect (DISC) command, followed by a Disconnect Mode (DM) response.

10.2.8 TEI Management

The LAPD for point-to-point connections in a terminal (Fig. 10.2-1) stores a TEI and checks the TEI in the address field of received frames to determine whether the frame is intended for it. It also places its TEI in the address fields of its outgoing frames. A terminal is "fixed" or "portable," depending on the manner in which its TEI value is entered.

A *fixed TE* is intended to be associated with a DSL on a long-term basis, say, several years. These terminals have a number of switches whose positions determine the TEI value. The switches are set by the person who installs the TE, and the settings remain unchanged as long as the TE remains on the DSL. Fixed TEs can have TEI values in the range 0–63.

A *portable TE* is intended to be moved from DSL to DSL. For example, a user may have a DSL at home and another DSL at the office and can take the TE along when moving between these locations. It is inconvenient to change the TEI value manually on each move, and portable TEs are therefore arranged to receive a TEI value (in the range 64–126) that is assigned by the exchange. The assignment of a TEI value involves "TEI management" messages between the portable TE and the exchange.

TEI management messages are carried in UI frames with TEI = 127 (broadcast TEI) and SAPI = 63 (network-layer management entity). The messages are:

ID-request. This is sent by a portable TE, to request the assignment of a TEI value.

ID-assigned. The response by the network to an ID-request. It includes the assigned TEI value.

ID-denied. This is the response by the network, denying an ID-request.

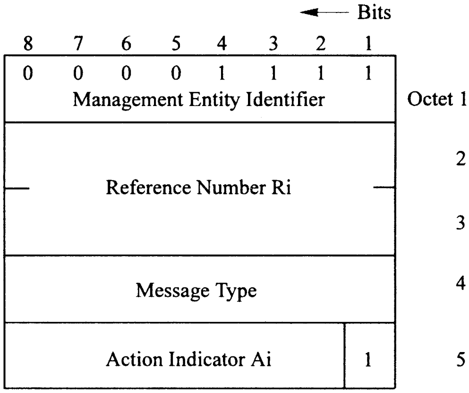


Figure 10.2-5. Information field of unnumbered information frame with a TEI management message. (From Rec. Q.921. Courtesy of ITU-T.)

ID-check-request. This is a command from the network to check out an assigned TEI value.

ID-check-response. This is the response by a portable TE to an ID-check-request.

ID-remove. This is a command sent from the network, to remove a portable TE with a specified TEI value from the DSL.

The information field of the UI frames is shown in Fig. 10.2-5. The code in octet 1 indicates a TEI management message. The message type code is in octet 4 (see Table 10.2-2). The message includes the parameters *Ri* (reference number) and *Ai* (action indicator).

We now describe the procedures for TEI assignment and TEI checking.

The *TEI assignment* procedure is shown in Fig. 10.2-6(a). When a portable TE is plugged into a DSL, it automatically sends an ID-request. Since the TE does not have a TEI, it generates a random reference number (Ri) to identify itself. A TE can request the assignment of a particular TEI value by specifying it in the Ai field or can leave the choice to the network by setting Ai = 127.

For each attached DSL, the network (local exchange) maintains a list of portable TEI values (range 64–126) that are currently assigned. When receiving an

TABLE 10.2-2 Message Type Codes in TEI Management Messages

Message Name	Message Type Code
ID-request	0000 0001
ID-assigned	0000 0010
ID-denied	0000 0011
ID-check-request	0000 0100
ID-check -response	0000 0101
ID-remove	0000 1010

Source: Rec. Q.921. Courtesy of ITU-T.

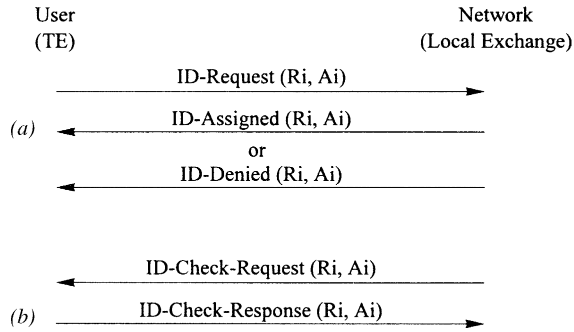


Figure 10.2-6. TEI management procedures: (*a*) TEI assignment and (*b*) TEI checking. All TEI management message are sent in UI frames with TEI = 127 and SAPI = 63. (From Rec. Q.921. Courtesy of ITU-T.)

ID-request on a DSL, the exchange consults the list. When it can assign a TEI, it broadcasts an ID-assigned message in which the Ri value is copied from the ID-request, and the assigned TEI value is in Ai.

All TEs on the DSL examine the message, but only the TE that has sent the request, and recognizes its Ri value, accepts the assigned TEI. Therefore, two or more TEs on a DSL can make simultaneous ID-requests without causing problems.

If the exchange cannot satisfy the ID-request because the requested TEI is already on the list of assigned TEIs for the DSL, or because all TEIs in the range 64–126 have already been assigned, it broadcasts an ID-denied message, again copying the Ri value from the received request. The TE then alerts its user that its request for a TEI assignment has been denied.

The *TEI checking* procedure allows the exchange to audit its list of assigned portable TEIs on a DSL—see Fig. 10.2-6(*b*). The exchange broadcasts an ID-check-request to the TEs on the DSL, in which Ai indicates the TEI value being checked, and the Ri field is set to 0. The network also starts a 200-ms timer. When a TE on the DSL has a TEI value that matches Ai, it respond with an ID-check-response that includes a randomly chosen number Ri and the received Ai value.

Under normal conditions, the network receives one ID-check-response before the timer expires. This indicates that there is one TE with the particular TEI value. If the timer expires and no response has been received, the network repeats the ID-check-request and restarts the timer. If the timer expires again before a response has been received, the network assumes that the TEI is no longer in use, deletes it from the list of assigned TEIs for the DSL, and makes a report for the maintenance staff of the exchange.

When the network receives more than one response to an ID-check-request, it knows that the same TEI value has mistakenly been assigned to more than one TE on the DSL. In this case, it broadcasts an ID-remove command, with the TEI value in Ai. The TEs whose TEI values match Ai then stop sending and accepting frames on the DSL and alert their users.

10.3 Q.931 CALL-CONTROL MESSAGES

This section describes the network-layer messages and parameters for the control of circuit-mode ISDN calls that have been specified in ITU-T. Recommendation Q.931 [8] and are generally known as Q.931 messages and parameters.

The ITU-T specification is an "umbrella": the messages and parameters used in individual countries are subsets of those defined by ITU-T. Some national Q.931 versions also include special parameter values and codes to support country-specific ISDN services and TE characteristics. The U.S. version of Q.931 has been specified by Bellcore (now Telcordia) [9].

Q.931 messages are located in the information fields of I and UI frames (Fig. 10.2-2).

10.3.1 Message Direction and Scope

We can distinguish Q.931 messages by their direction and scope—see Fig. 10.3-1. *Network-to-user* messages (a,c) are sent from a local exchange to a TE, and *user-to-network* messages (b,d) are sent in the opposite direction. In addition, we speak of *local* messages and *global* messages. A local message (a,b) is of significance only to the TE that sends or receives the message, and its local exchange. A global message (c) is a message, sent by a TE, that has significance for its local exchange, and for the distant TE. Global messages are transferred across the network and delivered (d) to the distant TE.

10.3.2 Q.931 Message Types

This section outlines the functions of the most important Q.931 messages and introduces the message acronyms that will be used throughout this chapter. Most message types can be sent from the user to the network and vice versa.

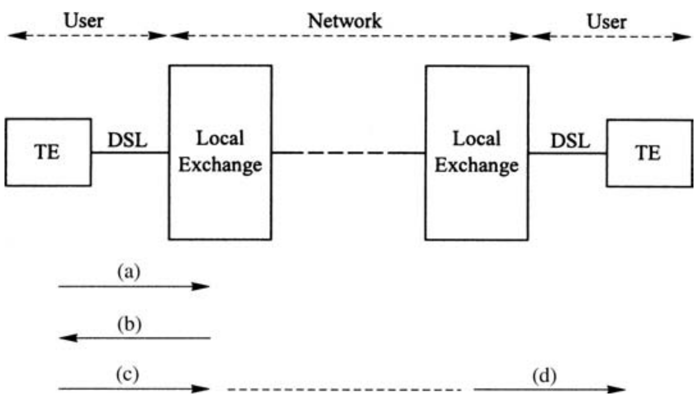


Figure 10.3-1. Classification of Q.931 messages by direction and scope.

Setup Message (SETUP). This is a global message that initiates a call. It is sent from the calling user to the network and from the network to the called user. It contains the called number and other information for the call setup.

Setup Acknowledgment Message (SETACK). This is a local message from the network to the calling user. It indicates that more address information is needed to setup the call.

Call Proceeding Message (CALPRC). This is a local message, sent from the network to the calling user or from the called user to the network. It confirms the receipt of a SETUP message and indicates that the complete address information for the call has been received, and that the setup is proceeding.

Progress Message (PROG). This is a local message, sent from the network to the calling user. It contains information about the progress of the call setup.

Alerting Message (ALERT). This is a global message, sent from the called user to the network and from the network to the calling user. It indicates that the called party is being informed (alerted) about the arrival of an incoming call.

Connect Message (CONN). This is a global message, sent from a called user to the network and from the network to a calling user. It indicates that the called user has answered.

Connect Acknowledgment Message (CONACK). This is a local message, acknowledging the receipt of a connect message.

Disconnect Message (DISC). This is a global message, sent from the user to the network and from the network to the user. It requests the release the connection.

Release Message (RLSE). This is a local message, acknowledging the receipt of a DISC message, and indicating that the sender has cleared the connection at its end.

Release Complete Message (RLCOM). This is a local message, acknowledging the receipt of a RLSE message and indicating that the sender has released the connection at its end.

Information Message (INFO). This is a global message. It is sent by a calling user, who enters called numbers from a keypad at her terminal, and contains one or more digits of the number. An INFO message can also be sent from the network to a calling user. In this case, it orders the user's TE to generate an audible progress tone (busy-tone, etc.).

10.3.3 General Message Format

The general format of Q.931 messages is shown in Fig. 10.3-2. The bits are numbered from right to left, and the first bit sent is bit 1 of octet a. All messages begin with a standard header that consists of the following three parts.

Protocol Discriminator. The message on the D-channel can belong to protocols other than Q.931 (e.g., the X.25 protocol for packet-mode communications). The code shown indicates the Q.931 protocol.

Call Reference Value. The *call reference value* (CRV) is an integer that identifies the call to which the message relates. Q.931 messages are call-related instead of trunk-related because the Q.931 protocol covers both circuit-mode and packet-mode calls. In a packet-mode call, there are no dedicated circuits (trunks), and the Q.931 messages for these calls are therefore call-related. For uniformity, Q.931 also uses call-related messages for circuit-mode calls.

At a TE, a call is identified uniquely by a CRV. At a local exchange, the call is identified uniquely by the combination of a CRV, TEI, and the identity of the DSL.

Octet b indicates the length (number of octets) of the CRV. On basic rate DSLs, CRV values range from 1 to 127, and the CRV is located in bits 7 to 1 of octet c—see Fig. 10.3-2. On primary rate DSLs, CRV values range from 0 through $2^{15} - 1$, and the CRV occupies two octets.

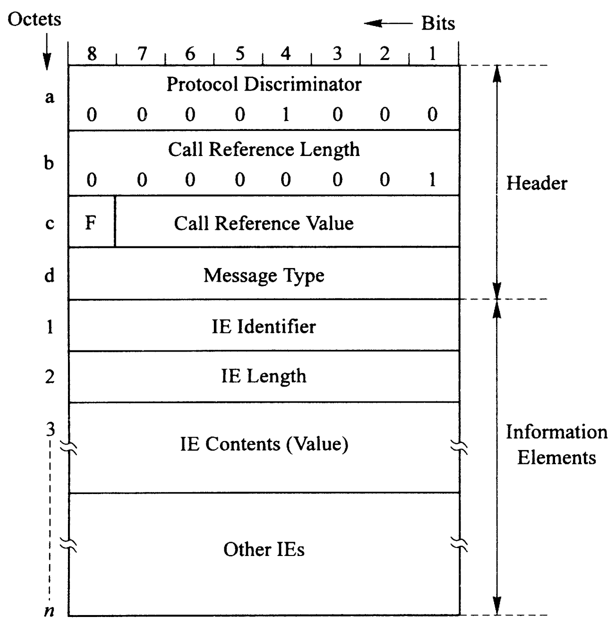


Figure 10.3-2. General format of Q.931 messages. (From Rec. Q.931. Courtesy of ITU-T.)

TABLE 10.3-1 Q.931 Message Type Codes

		← Bits							
Message Name	Acronym	8	7	6	5	4	3	2	1
Alerting	ALERT	0	0	0	0	0	0	0	1
Call Proceeding	CALPRC	0	0	0	0	0	0	1	0
Connect	CONN	0	0	0	0	0	1	1	1
Connect Acknowledgment	CONACK	0	0	0	0	1	1	1	1
Disconnect	DISC	0	1	0	0	0	1	0	1
Information	INFO	0	1	1	1	1	0	1	1
Progress	PROG	0	0	0	0	0	0	1	1
Release	RLSE	0	1	0	0	1	1	0	1
Release Complete	RLCOM	0	1	0	1	1	0	1	0
Setup	SETUP	0	0	0	0	0	1	0	1
Setup Acknowledgment	SETACK	0	0	0	0	1	1	0	1

Source: Rec. Q.931. Courtesy of ITU-T.

Flag bit (F) indicates whether the CRV has been assigned by the sender or the recipient of the message.

Message Type. Octet d identifies the message type. The coding is shown in Table 10.3-1.

In addition, a message includes a number of parameters, which are known in Q.931 documents as *information elements* (IEs).

10.3.4 Information Elements

An information element consists of three fields (Fig. 10.3-2):

IE Identifier. A one-octet field that holds the name of the IE—see Table 10.3-2.

IE Length. A one-octet field that indicates the length (number of octets) of the contents field.

IE Contents (or IE Value). A variable-length field that holds the actual information of the IE.

The identifier and length fields are in octets 1 and 2 of the IE. The contents field starts at octet 3.

In a message of a given type, the included IEs depend on the message direction. Moreover, an IE can be *mandatory* (always included in the message) or *optional* (included only when necessary).

The most important mandatory (M) and optional (O) IEs in user-to-network and network-to-user messages are listed in Tables 10.3-3 and 10.3-4 [8,9]. Each IE in the tables has a reference number (e.g., IE.1). We shall use these numbers when referring to IEs in later sections of this chapter.

TABLE 10.3-2 Information Element Identifiers

		← Bits							
Reference Number	Information Element	8	7	6	5	4	3	2	1
IE.1	Bearer capability	0	0	0	0	0	1	0	0
IE.2	Called party number	0	1	1	1	0	0	0	0
IE.3	Calling party number	0	1	1	0	1	1	0	0
IE.4	Called party subaddress	0	1	1	1	0	0	0	1
IE.5	Calling party subaddress	0	1	1	0	1	1	0	1
IE.6	Cause	0	0	0	0	1	0	0	0
IE.7	Channel ID	0	0	0	1	1	0	0	0
IE.8	High-layer compatibility	0	1	1	1	1	1	0	1
IE.9	Keypad	0	0	1	0	1	1	0	0
IE.10	Low-layer compatibility	0	1	1	1	1	1	0	0
IE.11	Progress indicator	0	0	0	1	1	1	1	0
IE.12	Signal	0	0	1	1	0	1	0	0
IE.13	Transit network selection	0	1	1	1	1	0	0	0
IE.14	User–user	0	1	1	1	1	1	1	0

10.3.5 Information Element Contents

This section describes the contents fields of the IEs listed in Table 10.3-2. At this point, it is sufficient to read quickly through these descriptions, which are primarily intended as reference material for Sections 10.4 and 10.5.

General Remarks. Q.931 coding uses the concept of "extended" octets. In some octets of contents fields, bit 8 is an extension bit (*ext*). If $ext = 1$, the octet consists of one octet. An "extended" octet is an "octet" that is extended to a next octet. This is indicated by $ext = 0$. For example, when the data item in octet N has a length of up to seven bits, octet N is not extended ($ext = 1$). Next, suppose that the data item in octet N has a length of 18 bits. The initial seven bits are then in octet N, which is marked as extended ($ext = 0$). Bits 8–14 are in the next octet, which is numbered N_a (first extension of octet N) and is also marked as extended. Finally, bits 15–18 are in an octet, which is marked as "not extended" ($ext = 1$).

The parameter *coding standard* appears in the contents field of some IEs (e.g., see Fig. 10.3-3). It indicates whether the field is coded according to ITU-T standards:

0	0	ITU-T standard
1	0	National standard
1	1	Network specific standard

This allows other standards organizations to define country-specific parameter codings. The descriptions that follow assume the ITU-T standard.

Some contents fields contain a fairly large number of data items. Only the most essential ones are discussed here (for additional information, see [8] and [9]). In the figures that follow, the blank fields represent data items that are not discussed.

We now examine the IE contents.

TABLE 10.3-3 Information Elements in Q.931 Call-Control Messages in the Direction User Network

		Message Acronym								
Reference Number	Information Element	ALERT	CALPRC	CONN	CONACK	DISC	INFO	RLSE	RLCOM	SETUP
IE.1	Bearer capability									M
IE.2	Called party number									M
IE.3	Calling party number									O
IE.4	Called party subaddress									O
IE.5	Calling party subaddress									O
IE.6	Cause					M		O	O	
IE.7	Channel identification	O	M	O						O
IE.8	High-layer compatibility									O
IE.9	Keypad						M			O
IE.10	Low-layer compatibility									O
IE.13	Transit network selection									O
IE.14	User−user	O		O		O		O	O	O

M, Mandatory; O, optional.

TABLE 10.3-4 Information Elements in Q.931 Call-Control Messages in the Direction Network User

		Message Acronym										
Reference Number	Information Element	ALERT	CALPRC	CONN	CONACK	DISC	INFO	PROG	RLSE	RLCOM	SETUP	SETACK
IE.1	Bearer capability										M	
IE.2	Called party number										M	
IE.3	Calling party number										O	
IE.4	Called party subaddress										O	
IE.5	Calling party subaddress										O	
IE.6	Cause					M		O	O	O		
IE.7	Channel identification		O								M	M
IE.8	High-layer compatibility										O	
IE.10	Low-layer compatibility										O	
IE.11	Progress indicator	O		O				M			O	O
IE.12	Signal	M		O	O	O	O	O	O	O	M	O
IE.14	User−user	O		O		O			O	O	O	

M, Mandatory; O, optional.

Bits 8	7	6	5	4	3	2	1	Octets
1 Ext	Coding Standard		Information Transfer Capability					3
0/1 Ext	Transfer mode		Information Transfer Rate					4
0/1 Ext			User Information Layer 1 Protocol					5
0/1 Ext			User Rate					5a

Figure 10.3-3. Format of IE.1: bearer capability. (From Rec. Q.931. Courtesy of ITU-T.)

IE.1—Bearer Capability (Fig. 10.3-3). This indicates the bearer service requested by the calling user.

Information transfer capability:
- 00000 Speech
- 01000 Unrestricted digital information
- 01001 Restricted digital information
- 10000 3.1-kHz Audio

Transfer mode:
- 00 Circuit mode
- 10 Packet mode

Information transfer rate:
- 00000 Packet mode
- 10000 64-kb/s Circuit mode

User information layer 1 protocol:
- 00001 ITU-T standardized rate adaptation
- 00010 Mu-law coded analog signal (Section 1.5.1)
- 00011 A-law coded analog signal

User rate (included only when the user information layer 1 protocol is 00001):
- 01111 Rate adaptation between 56 and 64 kb/s

IE.2—Called Party Number (Fig. 10.3-4). Octet 3a is not included.

Type of number:
- 001 International number
- 010 National number
- 100 Subscriber (directory) number
- 000 Special number

<table>
<tr><th colspan="8">Bits</th><th></th></tr>
<tr><th>8</th><th>7</th><th>6</th><th>5</th><th>4</th><th>3</th><th>2</th><th>1</th><th>Octets</th></tr>
<tr><td>0/1
Ext</td><td colspan="3">Type of Number</td><td colspan="4">Numbering Plan Identification</td><td>3</td></tr>
<tr><td>1
Ext</td><td colspan="3">Calling Number
Presentation</td><td colspan="4">Calling Number
Screening</td><td>3a</td></tr>
<tr><td>0</td><td colspan="7">Number Digits
(IA5 Characters)</td><td>4, etc.</td></tr>
</table>

Figure 10.3-4. Formats of IE.2 (called party number) and IE.3 (calling party number) (From Rec. Q.931. Courtesy of ITU-T.)

Numbering plan identification:

0001 ISDN/telephony numbering plan

Number digits: Each digit is coded as a seven-bit IA5 (International Alphabet No. 5) character and occupies one octet. Bits 7,6,5 are set to 001, and bits 4 to 1 represent the digit values from 0000 (zero) through 1001 (nine).

IE.3—Calling Party Number (Fig. 10.3-4). The coding of this IE is the same as for the called number, except that octet 3a is included.

Calling number presentation:

000 Presentation allowed (by calling party)

010 Presentation not allowed

Calling number screening:

0000 User-provided number, not screened by network (local exchange)

0001 User-provided number, has passed screening

0010 User-provided number, failed screening

0011 Number provided by the network

The screening status is important in calls to TEs that are not attended by humans (computers, facsimile machines). These TEs can be set up to accept only calls in which the calling number is as expected and has passed screening.

IE.4 and IE.5—Called and Calling Party Subaddress (Fig. 10.3-5). The IE holds a subaddress associated with a specific TE on the called or calling user's DSL and provides information for the selection of a compatible called TE. The IEs are transferred transparently (without inspection by the network) from the calling to the called user.

Type of subaddress:

000 Coded per ITU-T Recommendation X.213

010 Coding specified by the user

Bits 8	7	6	5	4	3	2	1	Octets
1 Ext	Type of Subaddress			Odd/even Indicator	0	0 Spare	0	3
Subaddress Information								4, etc.

Figure 10.3-5. Formats of IE.A (called party subaddress) and IE.5 (calling party subaddress). (From Rec. Q.931. Courtesy of ITU- T.)

Odd/even indicator: User-specified subaddress coding usually is a string of BCD code digits (two digits per octet).

0 Even number of digits

1 Odd number of digits.

IE.6—Cause (Fig. 10.3-6). Indicates why a setup has failed, or why a connection has to be released.

Location indicates where the message originated:

0000 (0) User
0010 (2) Public network serving the local user
0011 (3) Transit network
0100 (4) Public network serving the remote user
0111 (7) International network

The terms *local* and *remote* are as perceived by the recipient of the message. For example, in a message to user U, code (2) indicates that the message was originated in U's local network.

Cause value identifies the reason for the termination of a call or call setup. ITU-T has specified a large number of cause values [8], for example:

000 0001 (1) Unassigned number
000 0010 (2) No route to requested transit network
000 0011 (3) No route to destination
001 0000 (16) Normal call clearing
001 0001 (17) User busy
001 0010 (18) No user responding
001 0011 (19) User being alerted, no response
001 0101 (21) Call rejected
001 1010 (26) Clearing of nonselected user
001 1011 (27) Destination out of order
001 1100 (28) Invalid or incomplete called number
010 0010 (34) No trunk or B-channel available
010 1010 (42) Switching equipment congestion

011 1001	(57)	Requested bearer capability not authorized
100 0001	(65)	Bearer capability not implemented
111 0110	(102)	Recovery action on expiration of a timer

Bellcore (new Teliodia) has specified a number of additional causes (coded as network-specific) [9].

IE.7—Channel Identification (Fig. 10.3-7). This IE identifies the B-channel to be used for the call. Octets 3.1, 3.2, and 3.3 are included for primary rate DSLs only.

Interface type:

0 Basic rate DSL

1 Primary rate DSL

Information channel selection identifies the B-channel on a basic rate DSL:

01 Channel B_1

10 Channel B_2

Channel number identifies the B-channel on a primary rate DSL.

IE.8—High-Layer Compatibility (Fig. 10.3-8). This IE is used for the selection of a compatible TE on the called DSL. It is transferred transparently across the network.

<table>
<tr><th colspan="8">Bits</th><th></th></tr>
<tr><th>8</th><th>7</th><th>6</th><th>5</th><th>4</th><th>3</th><th>2</th><th>1</th><th>Octets</th></tr>
<tr><td>0/1
Ext</td><td colspan="2">Coding
Standard</td><td>0
Spare</td><td colspan="4">Location</td><td>3</td></tr>
<tr><td>1
Ext</td><td colspan="7">Cause Value</td><td>4</td></tr>
</table>

Figure 10.3-6. Format of IE.6 (cause). (From Rec. Q.931. Courtesy of ITU-T.)

<table>
<tr><th colspan="8">Bits</th><th></th></tr>
<tr><th>8</th><th>7</th><th>6</th><th>5</th><th>4</th><th>3</th><th>2</th><th>1</th><th>Octets</th></tr>
<tr><td rowspan="3">0</td><td colspan="7">Channel Identification</td><td rowspan="3">1</td></tr>
<tr><td>0</td><td>0</td><td>1</td><td>1</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td colspan="7">Information Element Identifier</td></tr>
<tr><td colspan="8">Length of Channel Identification Contents</td><td>2</td></tr>
<tr><td>1
Ext</td><td></td><td>Int.
Type</td><td>0
Spare</td><td></td><td></td><td colspan="2">Information Channel
Selection</td><td>3</td></tr>
<tr><td>1
Ext</td><td colspan="2">Coding Standard</td><td></td><td colspan="4"></td><td></td></tr>
<tr><td colspan="8">Channel Number</td><td></td></tr>
</table>

Figure 10.3-7. Format of IE.7 (channel identification). (From Rec. Q.931. Courtesy of ITU-T.)

8	7	6	5	4	3	2	1	Octets
1 Ext	Coding Standard		Interpretation			Presentation Method of Protocol Profile		3
0/1 Ext	High Layer Characteristics Identification							4

Figure 10.3-8. Format of IE.8 (high-layer compatibility). (From Rec. Q.931. Courtesy of ITU-T.)

High-layer characteristics identification:

000 0001 Telephony
000 0100 Facsimile group 4
110 0001 Teletex
110 0101 Telex

IE.9—Keypad Facility (Fig. 10.3-9). This IE holds one or more digits that are received from a keypad TE.

Keypad information: A string of digits coded as IA5 characters. Digits 0–9 are coded as in IE.2. Digits* and # are coded as 010 1010 and 010 0011.

IE.10—Low-Layer Compatibility. This IE is used for the selection of a compatible IE on the called DSL. It is transferred transparently across the network. The contents are the same as the contents of IE.1 bearer capability (IE.1).

IE.11—Progress Indicator (Fig. 10.3-10). This IE informs the user about certain characteristics of the connection or about tones or patterns on the B-channel.

Location holds the location of the IE originator and is coded as in IE.6 (cause).

Progress description:

000 0001 (1) Call is not end-to-end ISDN
000 0010 (2) Called equipment is non-ISDN
000 0011 (3) Calling equipment is non-ISDN
000 1000 (8) In-band tone or information now available on B-channel
000 1010 (10) Delay in response of called user

8	7	6	5	4	3	2	1	Octets
0	Keypad Information (IA5 Characters)							3, etc.

Figure 10.3-9. Format of IE.9 (keypad facility). (From Rec. Q.931. Courtesy of ITU-T.)

Bits 8	7	6	5	4	3	2	1	Octets
1 Ext	Coding Standard		0 Spare	Location				3
1 Ext	Progress Description							4

Figure 10.3-10. Format of IE.11 (progress indicator). (From Rec. Q.931. Courtesy of ITU-T.)

IE.12—Signal (Fig. 10.3-11). This IE allows the network to inform the user about in-band tones.

Signal value:

000 0000	(0)	Dial-tone on
000 0001	(1)	Ringing-tone on
000 0011	(3)	Recorder/congestion-tone on
000 0100	(4)	Busy-tone on
100 1111	(79)	Tones off

IE.13—Transit Network Selection (Fig. 10.3-12). Specifies the transit network selected for the call. If an ISDN user does not include this IE in the SETUP message for a long-distance or international call, the local exchange routes the call to the "normal" (default) interexchange carrier (IXC) of the user, which is stored at the exchange. A user can specify a different IXC for a call by including an IE.13 in the SETUP message.

Type of network identification:

010 National network identification

011 International network

Network identification plan:

0001 Carrier identification code

0011 Data network identification code

Network identification: A string of IA5 characters.

Bits 8	7	6	5	4	3	2	1	Octets
0	Signal Value							3

Figure 10.3-11. Format of IE.12 (signal). (From Rec. Q.931. Courtesy of ITU-T.)

Bits 8	7	6	5	4	3	2	1	Octets
1 Ext	Type of Network Identification			Network Identification Plan				3
0	Network Identification (IA5 characters)							4, etc.

Figure 10.3-12. Format of IE.13 (transit network selection). (From Rec. Q.931, Courtesy of ITU-T.)

IE.14—User–User (Fig. 10.3-13). User-to-user signaling is a service offered in several countries, but not in the United States. The IE, which can be included in SETUP, ALERT, CONN, DISC, RLSE, and/or RLCOM messages, allows the calling and called users to send up to 128 octets of information during the setup and release of a connection. The IE is transferred transparently across the network.

Protocol discriminator:

0000 0000 User-specific coding

0000 0010 Information in the form of IA5 characters

User information: This information is not standardized and has to be prearranged by individual pairs of ISDN users.

10.4 INTRODUCTION TO CALL-CONTROL SIGNALING

This section begins by outlining a typical sequence of Q.931 messages for a call between two users who are served by the same local exchange. It then introduces some ISDN call-control concepts and sets the stage for the additional examples of Section 10.5.

We shall use the message acronyms of Section 10.2.3 and refer to information elements by their reference numbers. When reading this section, it is helpful to look up the IE descriptions of Section 10.3.5.

Bits 8	7	6	5	4	3	2	1	Octets
Protocol Discriminator								3
User Information								4, etc.

Figure 10.3-13. Format of IE.14 (user–user). (From Rec. Q.931. Courtesy of ITU-T.)

10.4.1 Typical Signaling Sequence

We first examine the Q.931 messages in a typical signaling sequence for an intra-exchange call from user P to user Q (see Fig. 10.4-1). In this section, the focus is primarily on the messages, and the IEs are not discussed in detail. The message sequence is indicated by letters in parentheses, (a), (b), and so on.

Call Setup. User P starts the call by sending a SETUP message (a) from his terminal TE-P to the exchange. We assume that the message includes the complete called number (*en-bloc* signaling). The exchange then sends a SETUP message (b) on DSL-Q. It also assigns B-channels on DSL-P and DSL-Q and informs TE-P that the call setup has started, with a CALPRC message (c).

We also assume that TE-Q accepts the incoming call. It alerts the called user with a visual or audible signal and returns an ALERT message (d) to the exchange. The exchange then sends an ALERT message (e) to TE-P. If the bearer service for the call is "speech" or "3.1-kHz audio," the exchange also connects a ringing-tone source to the B-channel of the calling user (f).

When user Q answers, TE-Q sends a CONN message (g) to the exchange, which then connects the B-channels and sends a CONN message (h) to TE-P, indicating that the call has been answered. The exchange always acknowledges a received CONN message with a CONACK message (i). The figure also shows that TE-P acknowledges its CONN message with a CONACK message (j). This is not always the case: some TEs do not provide this message, and exchanges therefore do not require a CONACK response from a calling TE.

At this point, the users can begin to talk or to send data.

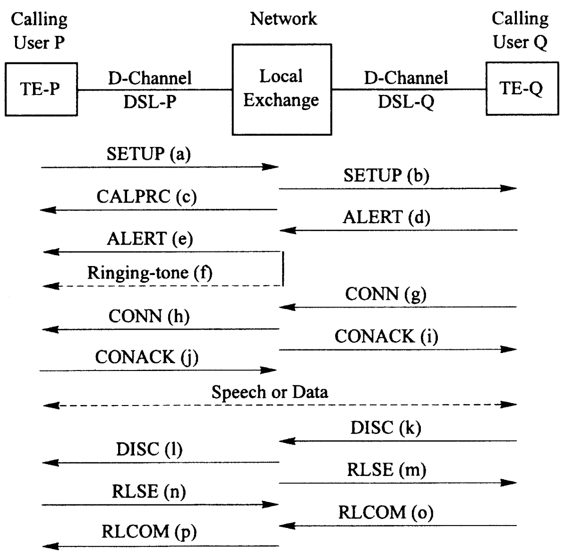


Figure 10.4-1. Signaling sequence for an intraexchange between ISDN users.

Call Clearing. In the signaling systems discussed so far, calls are cleared by the calling party. Q.931 follows the procedure of data communication networks, in which calls can be cleared by either party. In Fig. 10.4-1, called user Q initiates the clearing, by ordering TE-Q to send a DISC message (k). The exchange then sends a DISC message (1) to TE-P, clears its end of the B-channel to TE-Q, and sends a RLSE message (m) to that TE. Terminal TE-Q then releases the B-channel at its end and returns a RLCOM message (o). The B-channel is now available for a new call. The B-channel between TE-P and the exchange is released in a similar RLSE-RLCOM sequence.

10.4.2 Assignment and Release of Call Reference Values and B-Channels

We now explore how the exchange and the TEs inform each other about the assignment of *call reference values* (CRVs) and B-channels at the start of a call, and the release of these items when the call ends. We use the signaling sequence of Fig. 10.4-1.

The exchange, and each TE, maintains a pool of available CRVs and can assign a CRV to a call. The exchange knows the idle/busy status of the B-channels on its DSLs and assigns the channels to the call.

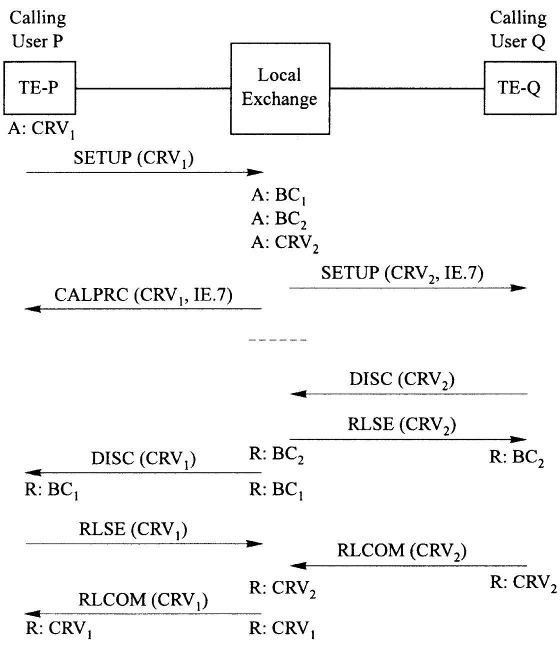


Figure 10.4-2. Assignment and release of call reference values and B-channels. IE.7, Channel identification (IE); CRV, call reference value; BC, B-channel; A, assigned; R, released.

Figure 10.4-2 shows only those messages in the sequence of Fig. 10.4-1 that play a role in informing the exchange and the TEs about assignments and releases of B-channels and CRVs.

Assignment of CRVs. The calling terminal TE-P assigns CRV_1 for its messages to and from the exchange and includes it in the SETUP message sent to the exchange. The exchange copies CRV_1 and includes it in all messages to TE-P.

The exchange selects CRV_2 for messages to and from the called terminal and includes it in its SETUP message to the called DSL-Q. If TE-Q accepts the call, it copies CRV_2 and uses it in its messages to the exchange.

Assignment of B-Channels. B-Channels BC_1 and BC_2, on the calling and called DSLs, are assigned by the exchange when it receives the SETUP message from the calling TE. To identify the assigned channels, the exchange includes an IE.7 (channel identification) in its CALPRC message to TE-P and in its SETUP message on the called DSL-Q.

Release of B-Channels and CRVs. At the end of the call, CRVs and B-channels have to be released at both ends. The release of BC_2 and CRV_2 is described first, because it occurs first in the signaling sequence of our example. When the exchange receives the DISC message from TE-Q, it releases BC_2 at its end and sends a RLSE message to TE-Q. In response, TE-Q releases BC_2 at its end, sends a RLCOM message to TE-P, and then discards CRV_2. When the exchange receives the RLCOM, it returns CRV_2 to its pool of CRVs. BC_2 and CRV_2 are now available for new calls.

The release of BC_1 and CRV_1 is done in a similar manner. When TE-P receives the DISC message, it releases BC_1 at its end and sends a RLSE message to the exchange. The exchange then releases the channel at its end, sends a RLCOM message to TE-P, and then discards CRV_1. On receipt of the RLCOM, TE-P returns CRV_1 to its pool of CRVs. BC_1 and CRV_1 are now available for new calls.

10.4.3 Called Terminal Selection

An ISDN user generally has TE of several types on a DSL. When making a call, the calling user has to include one or more "TE-selection" IEs in the SETUP message, to indicate the type of TE to which the caller wants to be connected. The selection of a called TE is described below [3,10].

Each TE on a DSL needs to examine all received SETUP messages, to determine whether its type is appropriate for the incoming call. The local exchange therefore sends SETUP messages in a frame with TEI = 127, which is the "broadcast" address to all TEs on a DSL.

The SETUP message sent includes all TE-selection IEs received from the calling user. Each TE stores one or more of these IEs, which are entered when the TE is installed. A TE compares the values of a received IE.n with the value of the corresponding stored IE.n. If there is a mismatch in any of the received and stored IE values, the TE disqualifies itself and does not respond to the SETUP message. The TE selection is thus an exclusion process and becomes sharper as the number of compared IEs increases.

The SETUP message to a called DSL always includes the requested bearer capability (IE.1), which is a TE-selection IE. ITU-T specifies three other selection criteria and IEs, which can be used singly or in combination. These are listed below.

Multiple Subscriber Numbers. At the request of a user, the telecom assigns several subscriber numbers to a DSL, and each TE stores one or more of these numbers (IE.2). In a call to a DSL with multiple numbers, the local exchange includes IE.2 in it broadcast SETUP message, and a terminal does not respond if the received IE.2 does not match any of its stored IE.2 values.

Subaddresses. In this case, the TEs on a DSL store a subaddress (IE.4), consisting of, say, some 2–4 digits. The calling party includes an IE.4 in his SETUP message. Only the TE whose stored IE.4 matches the value of the received IE.4 responds to the received SETUP message.

When a user has a group of TEs of the same type, multiple subscriber numbers or subaddresses have to be used to request a connection to a particular TE in that group.

Compatibility Checking. The TEs on a DSL can store high-level and/or low-level compatibility data. The calling user can include an IE.8 (high-level compatibility) and/or IE.10 (low-level compatibility) in the SETUP message. Again, only the TE or TEs whose stored data match the value(s) of the received IE(s) respond.

The subaddress, high-level compatibility and low-level compatibility IEs are optional parameters that are passed transparently (without examination by the exchange) from the calling to the called user.

10.4.4 Subscription Parameters

The local exchange stores a number of semipermanent parameter groups. These parameters define the services to which the user has subscribed and are used by the exchange when processing originating or terminating calls. This section examines the most important parameters that have been specified by Bellcore (now Telcordia) for local area (LATA) networks in the United States [9]. We can discern the following types of subscription-parameter groups.

DSL Subscription-Parameter Groups. These parameter groups are associated with the individual DSLs that are attached to the exchange and include the following parameters:

Type of DSL. This indicates the DSL type: basic access or primary rate access.

Bearer Capabilities. This is a list of bearer capabilities allowed on the DSL.

Directory Number (*DNs*). This is a list of directory (subscriber) numbers assigned to the DSL.

Calling Party Number Provision. This indicates whether the local exchange should accept a SETUP message from the DSL that does not include IE.3 (calling party number).

Calling Party Number Screening. This indicates whether the local exchange should screen a user-provided calling party number (IE.3) against the list of DNs assigned to the calling DSL.

Default Directory Number. This is the directory number to be sent to the called party when the calling party does not include IE.3 in the SETUP message.

DN Subscription-Parameter Groups. These groups are associated with the directory numbers (including the default DN) of the ISDN users served by the exchange. Examples of the parameters in these groups are:

Bearer Capabilities. This lists the bearer capabilities allowed for calls to and from this DN.

Interexchange Carrier Presubscription. This identifies the interexchange carrier, which should normally handle the user's long-distance calls. A calling user also can specify another carrier for a call, by including IE.13 (transit network selection) in his SETUP message.

Contention on Incoming Calls. This parameter has the values "allowed" and "not allowed" and is used by the terminating local exchange when several TEs on a DSL respond to a SETUP message—see Section 10.5.4.

Calling Number Delivery. This indicates whether the DN of the called user has calling line identification (or caller ID) service. If yes, the terminating local exchange includes IE.3 (calling party number) and IE.5 (calling party subaddress) in its SETUP broadcast message to the DSL.

Network-Provided In-band Tone/Announcement Service (*NPIBS*). On speech and 3.1-kHz audio calls, the calling users receive audible call-progress information (tones or announcements) when certain events occur during the setup of a call. NPIBS indicates whether this information is to be provided by the network (exchange) or by the TE.

Some TEs used in the United States can generate dial-, busy-, and reorder-tone, on command from the exchange. The directory numbers associated with these TEs are marked "not requiring NPIBS." Other TEs cannot generate these tones, and their directory numbers are marked as "requiring NPIBS."

Ringing-tone and recorded announcements are always provided by the network.

10.5 CALL-CONTROL EXAMPLES

We now examine a number of signaling sequences in more detail. When reading this section, it is helpful to look up the IE descriptions of Section 10.3.5.

10.5.1 Signaling Example

We start by revisiting the example of Fig. 10.4-1, this time including the most important information elements in the various messages (the IEs in Q.931 messages are listed in Tables 10.3-3 and 10.3-4). We also consider the influence of the subscription parameters (Section 10.4.4) on the call-control actions.

(a) SETUP Message from the Calling User. IE.1 (bearer capability) is mandatory. Since *en-bloc* address signaling has been assumed, IE.2 is included and contains the complete called number. IE.13 (transit network selection) is optional and is included only on long-distance calls if the calling user requests an interexchange carrier that is not the presubscribed carrier. IE.4, IE.5 (subaddresses), IE.8, and IE.10 (compatibility information), which play a role in the selection of the called terminal, are optional. IE.3 (calling party number) is mandatory or optional, depending on the contents of the "calling party number provision" subscription parameter of the calling DSL.

(b) SETUP Message to the Called User. Except for the calling and called numbers, the IEs present in the caller's SETUP message (a) are always included. IE.7 (identification of the assigned B-channel on the called DSL) is mandatory. If the subscription information of the called DSL lists more than one directory number, the exchange also includes IE.2 (called number). This selects a particular called TE on the DSL.IE.3 (calling number) is included if the "calling number delivery" subscription parameter of the called DN indicates that the user has this service, *and* IE.3 indicates that the caller allows his number to be presented. The calling number is either the number provided by the caller, and included in his SETUP message (a), or provided by the local exchange, which uses the default DN associated with caller's DSL.

(c) CALPRC Message to Calling User. IE.7 (identification of assigned-B channel on calling DSL) is mandatory.

(d) ALERT Message from Called User. No optional IEs are necessary in this example.

(e) ALERT Message to Calling User. IE.12 (signal), with value "ringing-tone on," is mandatory. On speech and 3.1-kHz audio calls, the exchange also sends ringing-tone on the B-channel and includes an IE.11 (progress indicator) in the message, to indicate that "in-band information is now available."

(g),(h) CONN Messages. Optional IE.12 (signal), indicating "tones off," is included in (h), because ALERT message (e) included an IE.12 that indicated "ringing-tone on."

(i),(j) CONACK Messages. Message (j) is sent only if the calling TE is able to provide it.

(k), (1) DISC Messages. IE.6 (cause) is mandatory. The general location is set to "user," and the cause value indicates "normal call clearing."

(m),(n) RLSE Messages. No optional IEs needed in this example.

(o),(p) RLCOM Messages. No optional IEs needed in this example.

10.5.2 Automatic Answering TEs

In the example of Fig. 10.4-1, the TE on the called DSL responds to a SETUP message with an ALERT message and follows up with a CONN message when the TE user answers. This is not always the case. The ALERT message is sent by a TE (e.g., a telephone) where a person is involved at the called end of the connection. *Automatic answering* TEs, such as facsimile machines, can receive calls without human intervention. When an automatic answering TE can accept an incoming call, it responds to the SETUP message with a CONN message.

10.5.3 Address Signaling

En-bloc Address Signaling. The example of Section 10.5.1 assumes *en-bloc* signaling of the called number. A TE arranged for *en-bloc* signaling waits until the user has entered the bearer service, the complete called number, and possibly other optional information elements, before sending the SETUP message. The procedures for entering the IEs into the TE are manufacturer-specific. When the local exchange receives the SETUP message, it assigns the B-channels to the call, sends a CALPRC message to the calling party, and starts the setup.

Overlap Address Signaling. Some TEs have a keypad and a handset, and the user signals the called address in the same way as on a dual-tone multifrequency (DTMF) telephone.

The SETUP messages from these TEs typically include only IE.1 (bearer capability), set to "speech" or "3.1-kHz audio." The local exchange assigns a B-channel and sends a SETACK message, which is a request to send digits—see Fig. 10.5-1. The message includes IE.7 (channel identification) and IE.12 (signal) set to "dial-tone on." If the subscription parameters indicate that the TE does not have NPIBS, the IE.12 is a command to generate dial-tone. If the TE is marked as having NPIBS, the exchange also sends dial-tone on the B-channel and includes an IE.11 (progress indicator), indicating "in-band information now available," in the message.

The calling TE then sends a number of INFO messages that include one or more IE.9s (keypad), each of which represents a digit of the called number. After receiving INFO message (1), the exchange returns INFO message (2), which has an IE.12 set to "tone off" and, if it has been sending dial-tone on the B-channel, it stops the tone. The user then completes the called number, which is transferred to the exchange in INFO messages (3) through (n). The exchange then starts to set up the connection.

10.5.4 Multiple TE Responses

When more than one TE on a called DSL responds to a SETUP broadcast message with a CALPRC, ALERT, or CONN message, the exchange has to award the call to one of them. The decision depends on whether contention on incoming calls is allowed (see Section 10.4.4). If contention is allowed, the exchange awards the call

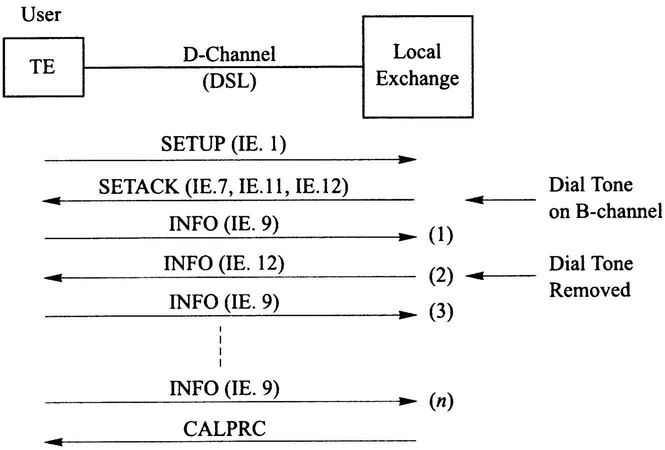


Figure 10.5-1. Overlap (keypad) signaling of called number.

to the first TE that sends a CONN message. If contention is not allowed, the call is awarded to the first TE that responds with a CALPRC, ALERT, or CONN message.

Figure 10.5-2 shows an example where contention is allowed, and TE-Q and TE-R have responded with an ALERT message. TE-R sends the first CONN message and is therefore awarded the call. The exchange sends a RLSE message to TE-Q that includes an IE.6 (cause) with location "public network of the local user" and cause value "clearing of nonselected user."

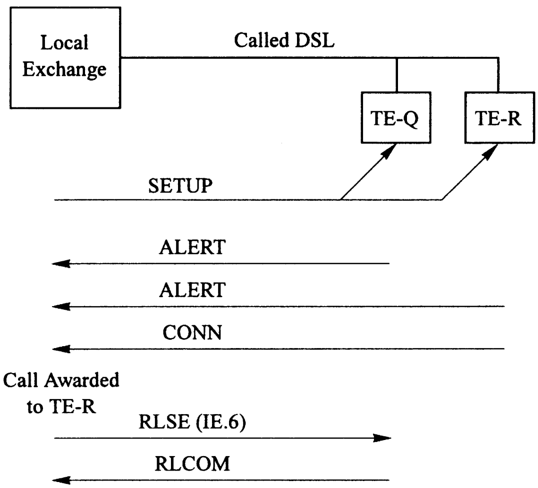


Figure 10.5-2. Multiple TE responding to SETUP message, contention allowed.

10.6 FAILED ISDN SETUPS

The subject of setup failures in ISDN is more complicated than in telephony. In the first place, there are more failure types: in addition to the failures (subscriber busy, invalid called number, etc.) that also occur in telephony, there are numerous failures that are specific to ISDN. In the second place, while a calling subscriber always receives an audible signal (tone or announcement) when a failure occurs, there are several ways in which a calling ISDN user can receive failure information.

This section explores a small number of failures that occur in the setup of intraexchange ISDN calls. An exhaustive treatment of ISDN setup failures in the U.S. networks can be found in [9].

10.6.1 Failure Causes

Failures can occur at several points in the setup (Fig. 10.6-1). We first consider failures that occur immediately after the exchange has received the caller's SETUP message. For example, the exchange can have determined that no B-channel is available on the calling DSL (F1), that the received called number is not assigned in the exchange (F2), or that no B-channel is available on the called DSL (F3). Failure F3 is known as "user interface busy."

Failures F4 and F5 occur after the exchange has sent its SETUP and CALPRC messages on the called and calling DSLs. After sending the message, the exchange starts a 4-s timer T_1 and, on expiration of the timer, evaluates the received responses. In failure F4, the exchange has received RLCOM messages only. This means that all TEs that satisfy the TE-selection criteria are busy or unable to accept the call for other reasons. Failure F4 is known as "user-determined busy."

Next, consider the case that no response has been received when T_1 expires. The SETUP broadcast message on the called DSL (Section 10.4.3) is sent in an UI (unnumbered information) frame (Section 10.2.5) that is not acknowledged by the TEs. It is possible that the message has incurred a transmission error and has been discarded by the TEs. The exchange therefore broadcasts a second SETUP, restarts T_1, and sends a PROG message to the calling user that includes an IE.11 (progress indicator) with value "delay in response of called user." On speech and 3.1-kHz audio calls, the exchange also starts to send ringing-tone on the calling B-channel, and the PROG message includes a second IE.11, with value "in-band information now available" (not shown in Fig. 10.6-1). If no response has been received when T_1 expires again, the exchange concludes that the TEs on the DSL are not responding (F5).

Finally, suppose that the exchange has received one or more ALERT responses from the called DSL. It then sends an ALERT message to the calling TE. On speech and 3.1-kHz audio calls, the exchange also returns ringing tone on the calling B-channel, and the ALERT message includes an IE.11 set to "in-band information now available." The exchange then starts a 5-minute timer (T_2). If the timer expires and no CONN message has been received, the exchange concludes that the users at the TEs are not answering (F6).

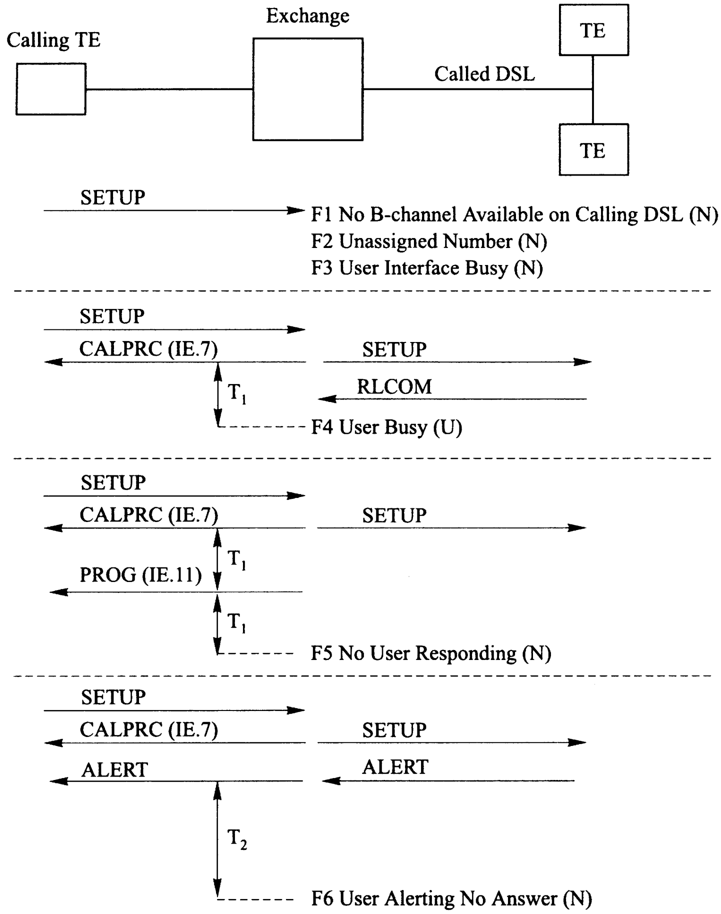


Figure 10.6-1. Setup failure examples. N, Failures determined by network (exchange); U, failures determined by user.

10.6.2 Treatments of the Calling TE

On failed setups, the exchange has to inform the calling user and to release B-channel connections that already have been established when the failure occurs.

The calling TE always receives a message that indicates the cause of the failure. On speech and 3.1-kHz audio calls, the user at the TE also receives in-band information (a tone or announcement) sent by the exchange on the B-channel or generated by the TE in response to a "signal" command (IE.12) from the exchange.

The exchange gives one of the following failure treatments to the calling TE, depending on whether the failure occurs before or after it has sent a CALPRC message for the call (which identifies the allocated B-channel to the TE), and on whether it has to provide in-band information:

	In-band Information: Not Provided	In-band Information: Provided
Failure before CALPRC	FT1	FT1-IB
Failure after CALPRC	FT2	FT2-IB

The provision of in-band information depends on several factors. On 64-kb/s calls, the exchange never provides in-band information. On speech or 3.1-kHz audio calls, the exchange always provides ringing-tone and announcements. However, busy-tone and reorder-tone are sent only if the calling TE has network-provided in-band information service (NPIBS) (Section 10.4.4). Obviously, no in-band information is sent when no B-channel is available on the calling DSL (Fl).

Table 10.6-1 lists the treatments for the failure cases described in the previous section. On speech and 3.1-kHz audio calls, the exchange is providing ringing-tone when failure F5 (no user responding) or F6 (user alerting, no answer) occurs. On failure F5, the exchange leaves the tone on and gives the FT2-IB treatment. On failure F6, the exchange turns the tone off and gives the FT2 treatment.

The treatments are outlined in Fig. 10.6-2. In FT1, the B-channel connection is not yet established, and the exchange ends the setup with a RLCOM message. In FT2, the exchange initiates the release of the B-channel with a DISC message. In treatments FT1-IB and FT2-IB, the exchange sends a PROG message, returns a tone or announcement on the B-channel, and then waits for a DISC message from the calling TE. In FT1-IB, the PROG message is preceded by a CALPRC message, because the B-channel connection, which is needed for the transfer of the in-band information, has not yet been established.

TABLE 10.6-1 Treatments of Calling TE on Setup Failures

			Bearer Services		
			64 kb/s	3.1 kHz	
				Calling TE	
Failure	Description	In-band Information		No NPIBS	NPIBS
F1	No B-channel available on calling DSL (N)	—	FT1	FT1	FT1
F2	Unassigned number (N)	Announcement	FT1	FT1-IB	FT1-IB
F3	User busy (N)	Busy-tone	FT1	FT1	FT1-IB
F4	User busy (U)	Busy-tone	FT2	FT2	FT2-IB
F5	No User responding (N)	Ringing-tone	FT2	FT2-IB	FT2-IB
F6	User alerting, no answer (N)	Ringing-tone Turned off	FT2	FT2	FT2

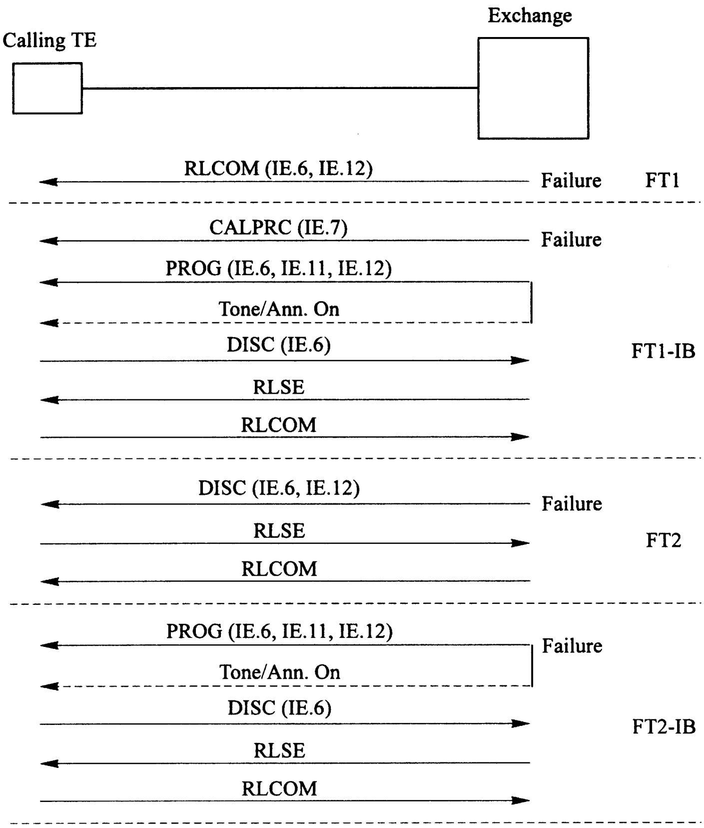


Figure 10.6-2. Failure treatments.

Table 10.6-2 shows the contents of IE.6 (cause) and IE.12 (signal) in the RLCOM, DISC, and PROG messages for the various failures. The contents generally do not depend on the bearer service or on whether the calling TE has NPIBS. However, in failure F2 (unassigned number), IE.12 is not included on speech or 3.1-kHz audio calls and is set to "reorder-tone" on 64-kb/s calls.

When the exchange is giving an FT1-IB or FT2-IB treatment, the PROG message also includes an IE.11 (progress indicator) with value "in-band information now available."

10.6.3 Treatment of Called TEs

In failure F6 (no answer), the exchange initiates the release of the B-channel on the called DSL with RLSE messages to the called TEs that have sent ALERT messages. The IE.6 (cause) in the message has location "public network serving local user" and value "recovery action on expiration of timer."

TABLE 10.6-2 Coding of IE.6 (Cause) and IE.12 (Signal) in Messages RLCOM, DISC, and PROG Sent in Setup Failures

		IE.6		
Failure	Description	Location	Cause Value	IE.12
F1	No B-channel available on calling DSL (N)	2	34	3
F2	Unassigned number (N)	2	1	3 (note 1)
F3	User busy (N)	4	17	4
F4	User busy (U)	0	17	4
F5	No user responding (N)	4	18	1
F6	User alerting, no answer (N)	4	19	63

Notes: (1) IE.12 included only when bearer service is 64 kb/s. (2) Functional meanings of the listed decimal codes:

IE.6—Location	
0	User
2	Public network serving the local user
4	Public network serving the remote user

IE.6—Cause Value	
1	Unassigned number
17	User busy
18	No user responding
19	User alerting, no answer
34	Calling B-channel not available

IE.12	
1	Ringing-tone on
3	Reorder-tone on
4	Busy-tone on
63	Tones off

The exchange does not need to release the B-channel in the other failures, because it has not broadcast a SETUP message (failures Fl, F2, and F3), or has sent a SETUP message and has received no response or “busy” responses only (F4).

10.7 ACRONYMS

Ai	Action indicator
ALERT	Alerting message
B-Channel	64-kb/s Channel for circuit-mode communications
BCD	Binary coded decimal digit
BSN	Backward sequence number
CALPRC	Call Proceeding message
CB	Check-bit
CONACK	Connect Acknowledgment message
CONN	Connect message

CRV	Call reference value
D-Channel	Signaling and packet channel on DSL
DISC	Disconnect message
DL	Data link
DM	Disconnect Mode frame
DN	Directory (subscriber) number
DSL	Digital subscriber line
DSLI	DSL identifier
DSS1	Digital Subscriber Signaling System No. 1
DTE	Data terminal equipment
ET	Exchange termination
Ext	Extension bit
FCS	Frame check sequence
FDM	Frequency-division multiplexing
FSN	Forward sequence number
I	Information frame
IA5	International alphanumeric coding
ID	Identity
IE	Information element
INFO	Information message
ISDN	Integrated Services Digital Network
ISUP	ISDN User Part
ITU	International Telecommunication Union
ITU-T	Telecommunication Standardization Sector of ITU
IXC	Intertexchange carrier
LAPD	Link Access Protocol for D-channel
LATA	Local access and transport area
M	Mandatory
MSU	Message signal unit
MTP	Message Transfer Part
NPIBS	Network-provided in-band information service
N(R)	Receive sequence number
N(S)	Send sequence number
NT	Network termination
O	Optional
PCM	Pulse code modulation
PROG	Progress message
REJ	Reject frame
Ri	Reference number
RLCOM	Release Complete message
RLSE	Release message
RNR	Receive Not Ready frame
RR	Receive Ready frame
SABME	Set Automatic Balanced Mode Extended frame

SAPI	Service access point indicator
SETACK	Setup Acknowledgment message
SETUP	Setup message
SS7	CCITT Signaling System No. 7
SU	Signal unit
TE	User terminal equipment
TEI	Terminal endpoint identifier
TUP	Telephone User Part
UA	Unnumbered Acknowledgment frame
UI	Unnumbered Information frame

10.8 REFERENCES

1. W. Stalling, *ISDN, An Introduction*, Macmillan, New York, 1989.
2. P. K. Verma (ed.), *ISDN Systems, Architecture, Technology and Applications*, Prentice Hall, Englewood Cliffs, NJ, 1990.
3. J. M. Griffiths, *ISDN Explained*, John Wiley & Sons, Hoboken, NJ, 1990.
4. Roger L. Freeman. *Telecommunication System Engineering*, Wiley–Interscience, Hoboken, NJ, 1996.
5. *Data Communication Networks: Services and Facilities, Interfaces*, Rec. X.1–X.32, ITU-T, Geneva, 1989.
6. R. J. Manterfield, *Common-Channel Signalling*, Peter Peregrinus Ltd., London, 1991.
7. *Digital Subscriber Signalling System No. I (DSS1), Data Link Layer*, Rec. Q.920–Q.921, ITU-T, Geneva, 1989.
8. *Digital Subscriber Signalling System No. 1 (DSS1), Network layer, User-Network Management*, Rec. Q.930–Q.940, ITU-T, Geneva 1989.
9. *ISDN Access Call Control Switching and Signalling Requirements*, Report TR-TSY-000268, Issue 3 (1989), Bellcore, Piscataway, NJ.
10. D. R. Davies and R. Haslam, Developments in ISDN: Signalling Standards and Terminals, *Br. Telecommun. Technol J.*, **9**, No. 2, 1991.

11

ISDN USER PART

This chapter resumes the discussion of SS7 signaling and describes the ISDN (Integrated Services Digital Network) user part of SS7.

11.1 INTRODUCTION

The ISDN User Part (ISUP) has been designed for use in the circuit-switched part of ISDNs (see Fig. 10.1-2). It includes messages and procedures for the control of inter-exchange calls between two (analog) subscribers, two ISDN users, and an ISDN user and a subscriber [1,2].

As in TUP signaling, ISUP call-control signaling is primarily link-by-link, but also includes procedures for end-to-end signaling (Section 11.5).

Telecoms that have installed TUP signaling in their networks have mostly converted to ISUP signaling. The telecoms in the United States chose to bypass TUP, moving directly from multifrequency and CCIS (Chapters 4 and 5) to ISUP.

11.1.1 ISUP Call-Control Requirements

The major difference between TUP and ISUP is that ISUP also supports signaling for calls in which one or both parties are ISDN users. ISUP has been designed to meet the signaling requirements for ISDN-related services.

Support of Bearer Services. The bearer service requested by a calling user can impose restrictions on the network equipment that can be used for the call. These limitations are listed in Table 11.1-1. For example, speech calls from analog lines can be carried on analog or digital (PCM) trunks that may be equipped with circuit multiplication equipment and on which echo controls may be enabled

Signaling in Telecommunication Networks, Second Edition, by John G. van Bosse and Fabrizio U. Devetak

TABLE 11.1-1 Trunk, Transmission, Exchange, and Signaling Types Allowed in Connections for Individual Bearer Services

	Speech	Bearer Service 3.1-kHz Audio	64-kb/s Data
Analog (FDM) trunk group	A	A	N
64-kb/s Digital trunk group	A	A	A
Trunk group with ISUP signaling	A	A	A
Trunk group with other signaling	A	A	N
Circuit multiplication equipment	A	N	N
Echo control	A	N	N
Exchange with analog switchblock	A	A	N
Exchange with digital switchblock	A	A	A

A, Allowed; N, not allowed.

(Section 1.6). Also, speech calls can be routed via exchanges with analog or digital (64-kb/s) switchblocks (Section 1.7).

On the other hand, ISDN (64-kb/s) calls require digital trunks and have to be routed via exchanges with digital switchblocks. Moreover, the bit patterns in the transmitted octets cannot be changed. Therefore, circuit multiplication equipment is not allowed, and echo cancelers have to be disabled.

Exchanges have to take these restrictions into account when selecting an outgoing trunk for an ISDN call.

Transfer of Q.931 Information. In interexchange calls, the calling and called users should be able to send and receive the same Q.931 messages and information elements as in intraexchange calls (Sections 10.3 and 10.6). ISUP signaling therefore includes messages and parameters to transfer the Q.931 information across the network. For example, the information elements included in a Q.931 SETUP message from the calling user for the selection of a particular called TE (Section 10.4.3) are transferred in ISUP messages to the terminating local exchange, which then includes them in its Q.931 SETUP message to the called DSL. Also, the "cause" information element of Q.931, which indicates the reason why a connection should be released, is transferred by ISUP messages to the calling or called local exchange, which then informs the ISDN user.

Services for Analog Subscribers. In present telecommunication networks, the number of digital subscriber lines is still a small fraction of the total number of lines. ISUP signaling includes messages and procedures to support the services already available to analog subscribers. In countries that are using TUP signaling, ISUP supports the services available with TUP (Section 9.4).

In the United States, ISUP supports a number of services for business customers and all services that require the transfer of the calling party number to the called party (Section 3.5.2).

11.1.2 ISUP Versions

ITU-T has defined ISUP in the Q.761–Q.764 recommendations [3–6], which are the basis for the descriptions of ISUP in Sections 11.2 through 11.7.

ISUP is a very powerful and complex signaling system. Implemented versions usually include only a subset of the above recommendations. On the other hand, country-specific versions of ISUP—notably the U.S. version—include signals and procedures that have not been defined by ITU-T. Examples of implemented ISUP versions are given in Sections 11.8 and 11.9.

11.1.3 Interface with MTP

The interface between ISUP and the message transfer part (MTP) at a signaling point is as shown in Fig. 8.7-1. The MTP-transfer primitives include three groups of parameters:

Service Indicator Octet (*SIO*). This group consists of parameters SI (service information) and SSF (subservice field). In ISUP messages, SI = 0101.

Routing Label (*RL*). This group consists of DPC (destination point code), OPC (originating point code), and SLS (signaling link selector).

User Message (*UM*). This group holds the ISUP information that is passed transparently by MTP.

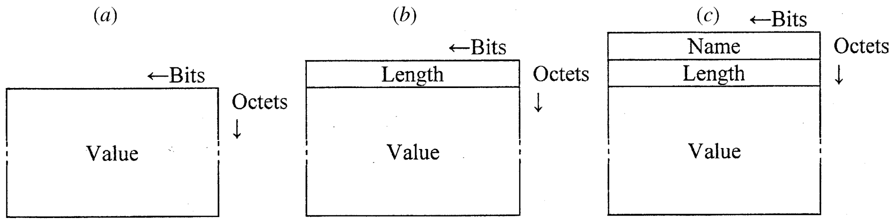
MTP-status, MTP-pause, and MTP-resume indications are sent by MTP to ISUP when a signaling route set to a particular destination becomes congested, unavailable, or available again.

11.1.4 Interface with SCCP

ISUP may also use the services of the Signaling Connection Control Part (SCCP) of SS7, discussed in Chapter 15. The interface with SCCP supports a method of end-to-end signaling between the originating and terminating exchange for a call (Section 11.5).

11.2 ISUP MESSAGES, FORMATS, AND PARAMETERS

This section describes the most important ISUP call-control messages. Since we have returned to SS7, we speak again of *parameters*. ISUP messages have *mandatory* parameters (always present in messages of a given type) and *optional* parameters (included only when necessary). The number of optional parameters and the number of messages that can include these parameters are much larger than in TUP. This gives ISUP messages the flexibility to accommodate ISDN signaling requirements. This flexibility also has resulted in a rather small number of call-control messages (about 10 in ISUP versus 30 in TUP).



(a): mandatory, fixed-length (MF) parameters. (b) mandatory, variable-length (MV) parameters. (c): optional (OP) parameters

Figure 11.2-1. ISUP parameters.

11.2.1 Parameter Fields and Types

ISUP parameters [5] consist of one, two, or three fields—see Fig. 11.2-1. The Name field (one octet) contains the parameter name and thus indicates the meaning of the parameter; the Length field (one octet) indicates the number of octets in the Value (or "contents") field, and this latter field contains the actual information.

We distinguish three parameter types:

Mandatory, Fixed Parameters (*MF Parameters*). In a particular message type, these parameters are always present, and their Value fields have fixed lengths.

Mandatory, Variable Parameters (*MV Parameters*). In a particular message type, these parameters are always present, but the lengths of their Value fields are variable.

Optional Parameters (*OPs*). These parameters, which may have fixed or variable length Value fields, appear in a given message type only when necessary.

11.2.2 ISUP Message Format

This section describes the general format of ISUP messages (Fig. 11.2-2). The *circuit identification code* (CIC) has 14 bits and is located in octets 1 and 2. In a network, an ISUP trunk is uniquely identified by the combination of DPC and OPC (located in RL—see Fig. 7.3-2) and of CIC. In TUP signaling, the four low-order bits of CIC are also used as the signaling link selector (SLS—Section 9.1.1). In ISUP signaling, SLS and CIC are separate parameters. This allows the rotation of the contents of SLS at the signaling points along the message route (see Example 3 of Section 8.8.5).

The *message type code* (octet 3) identifies a particular ISUP message. CIC and the message type code are mandatory parameters (Value fields only).

Octets 4 through n hold the message parameters. In this example, A through F are mandatory fixed (MF) parameters. In a particular ISUP message type, these parameters always appear—in predetermined order—starting at octet 4. Since their lengths are fixed, the locations of the MF parameters in the message are also fixed and known by the software entity that reads the message. Therefore, name and length fields are not needed.

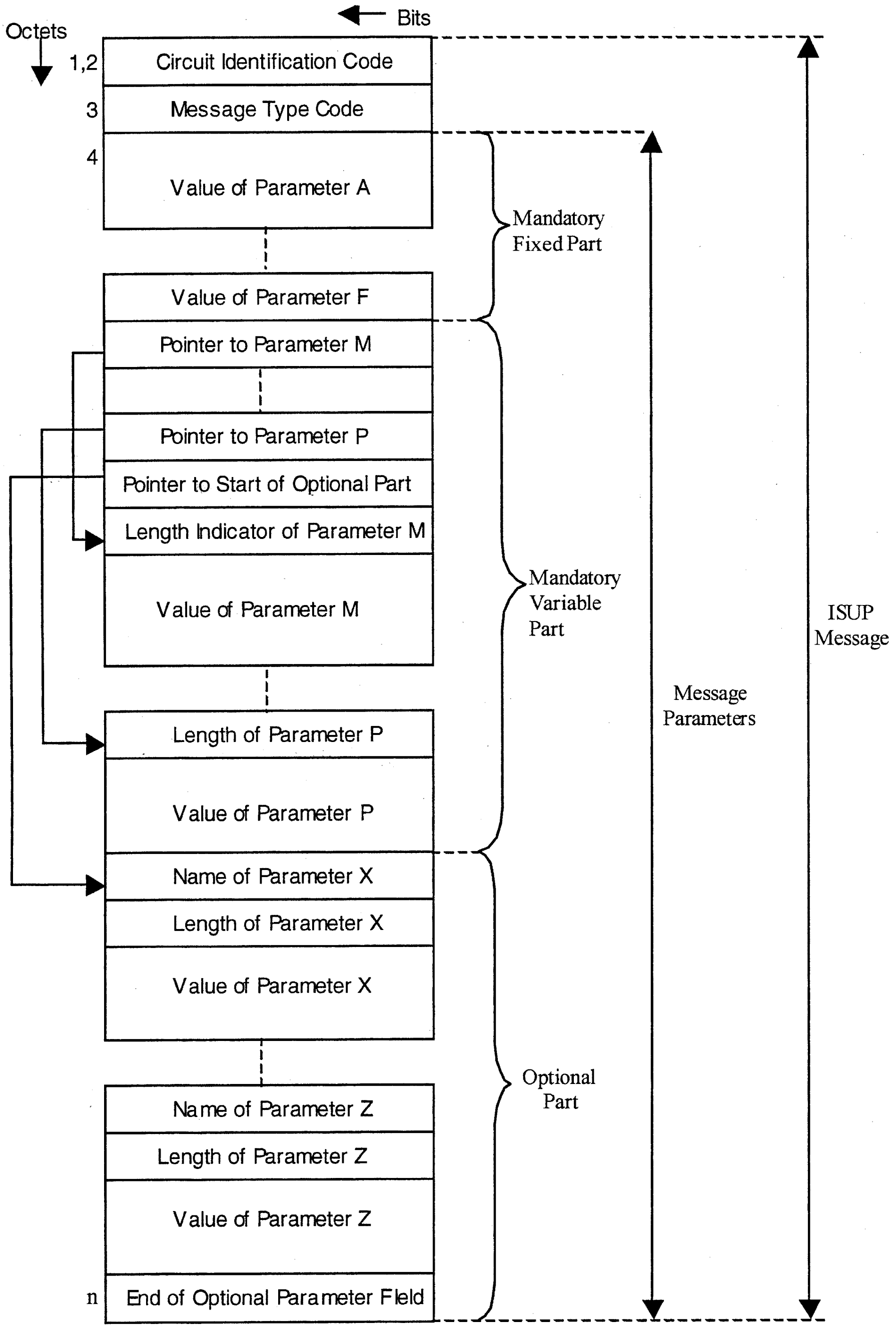


Figure 11.2-2. General ISUP message format. (From Rec. Q.763. Courtesy of ITU-T.)

The mandatory variable (MV) part of the message-parameter field begins with a block of one-octet pointers to the locations of the length fields of the MV parameters (in this example, M through P). In messages of a particular type, this block always starts at the same location, and the pointers appear in a predetermined order. The software entity that reads the message can thus determine the locations of any MV parameter. Name fields are therefore not necessary.

The octet following the pointer of the last MV parameter holds the pointer to the first octet of the optional part of the message-parameter field, where the OP parameters X through Z are located. Optional parameters can appear in any combination, and in any order, and therefore appear with their name, length, and value fields.

When a message does not include optional parameters, the pointer to the optional part of the message is coded 0000 0000. When a message includes optional parameters, an octet coded 0000 0000 follows the value field of the last parameter.

The message structure described above is also used for messages of SCCP.

11.2.3 Call-Control Messages

The most important ISUP call-control messages are outlined below. Most messages are the counterparts of TUP messages but generally include more parameters (Table 11.2-1).

Initial Address Message (IAM). As in TUP signaling, the IAM is the first message in a call setup. It includes the called number and other parameters that are of importance for the setup. In most countries, and in the international network, ISUP address signaling is *en-bloc* (the IAM includes the complete called number). However, ITU-T has also specified a Subsequent Address message (SAM)

TABLE 11.2-1 Message Type Codes of ISUP Call-Control Messages

Acronym	Name	Message Type Code
ACM	Address Complete	0000 0110
ANM	Answer	0000 1001
APM	Application Transport	0100 0001
CPG	Call Progress	0010 1100
CON	Connect	0000 0111
COT	Continuity	0000 0101
FOT	Forward Transfer	0000 1000
IAM	Initial Address	0000 0001
INF	Information	0000 0100
INR	Information Request	0000 0011
PAM	Pass-Along	0010 1000
REL	Release	0000 1100
RES	Resume	0000 1110
RLC	Release Complete	0001 0000
SUS	Suspend	0000 1101

Source: Rec. Q.763. Courtesy of ITU-T.

for overlap address signaling. The IAM then holds the initial digits of the called address, and the remaining digits are transferred in one or more SAMs.

Continuity Message (COT). This message is sent forward, during call setup, if the outgoing exchange has made a continuity test on an outgoing trunk. It indicates whether the test has been successful.

Address Complete Message (ACM). This is a backward message. When sent by the terminating local exchange, it indicates that the exchange is ringing the called subscriber or has received a Q.931 ALERT message from a terminal on the called digital subscriber line (DSL). When sent by an intermediate exchange, it indicates that the exchange has seized an outgoing trunk that does not have ISUP signaling.

Call Progress Message (CPG). This is a backward message sent by a terminating exchange that has already sent an ACM message, when it needs to report the occurrence of an "event" in the call setup.

Answer Message (ANM). This is a backward message sent when the called party answers.

Connect Message (CON). This message is sent on ISDN calls on receipt of a connect indication from the called party when ACM has not been sent. It is equivalent to ACM followed by ANM.

Release Message (REL). This is a general message that requests the immediate release of a connection. It can be sent forward or backward and includes—like the Q.931 DISC (Disconnect) message—a Cause parameter that indicates the reason for and the originator of the release. In other interexchange signaling systems, the release of a connection is normally initiated by the calling subscriber. In ISUP signaling, the calling and the called ISDN users can initiate a release.

A REL message is also sent by an intermediate exchange, or by the terminating local exchange, when the setup of a call in not possible.

Release Complete Message (RLC). This is a forward or backward message, sent by an exchange in response to a received REL message for a trunk. It indicates that the exchange has completed the release of the trunk at its end. The RLC message has the same function as the Release-Guard message of TUP signaling.

Suspend Message (SUS). This is a backward message that is similar to the Clear-Back message of TUP. It requests to suspend the call, but to leave the connection intact.

Resume Message (RES). This message requests to resume a suspended call.

Forward Transfer Message (FOT). An outgoing operator, who is assisting a subscriber or ISDN user in a call setup, can request the help of an incoming operator by sending a (forward) FOT message.

Information Request Message (INR). This is a request for additional call-related information, usually sent from the terminating exchange to the originating exchange of a connection (end-to-end signaling).

Information Message (INF). This is a response to an INR message and contains the requested information.

Pass-Along Message (PAM). This message holds another message that is transferred end-to-end with the pass-along method. This message and method is for national use only.

Application Transport Message (APM). This message is used by applications to exchange application-specific information by means of the *application transport mechanism* (Section 11.6.10).

11.2.4 Parameters in Call-Control Messages

This section describes the most important parameters of ISUP call-control messages [5]. At this point, it is suggested that the descriptions are merely perused and referred back to when reading later sections of this chapter.

To allow the reader to locate a parameter description quickly, each parameter has a reference number (e.g., Par.1). In the descriptions of this section, the parameters appear in the order of their reference numbers. In Sections 11.3–11.7, a parameter is always identified by its name and reference number.

The parameters, their reference numbers, and the messages in which they can appear are listed in Table 11.2-2.

In this section, the focus is on the parameter contents, and only their Value fields are shown in the figures. The parameter Name codes are listed in Table 11.2-3.

Par.1 Access Transport. This is an optional parameter in IAM that contains one or more Q.931 information elements. The IEs are transferred transparently between the ISDN users. The parameter can include IE.4 (called party subaddress), IE.5 (calling party subaddress), IE.8 (high-layer compatibility), IE.10 (low-layer compatibility), and IE.11 (progress indicator)—see Section 10.3.5.

Par.2 Application Transport (ATP). This is an optional parameter in ACM, ANM, APM, CON, CPG, and IAM messages (Fig. 11.2-3). It carries application-specific data (Section 11.6.10). Ext is the extension indicator (0 = last octet, 1 = more octets follow). ACI is the *application context identifier*, which identifies the application. SNI is the *send notification indicator* (0 = do not send, 1 = send). RCI is the *release call indicator* (0 = do not release, 1 = release). The SI and APM

TABLE 11.2-2 Parameters in ISUP Call-Control Messages

Reference Number	Parameter Name	Message Acronyms[a]													
		ACM	ANM	APM	CON	COT	CPG	FOT	IAM	INF	INR	REL	RES	RLC	SUS
Par.1	Access transport	O	O			O			O	O		O			
Par.2	Application transport (ATP)	O	O	O	O		O		O						
Par.3	Automatic congestion Level											O			
Par.4	Backward call indicators	M	M		M		O								
Par.5	Call reference	O			O		O	O	O	O	O	O	O		O
Par.6	Called party number								M						
Par.7	Calling party category								M	O					
Par.8	Calling party number								O	O					
Par.9	Cause indicators	O					O					M			
Par.10	Circuit assignment map								O						
Par.11	Closed user group interlock code								O						
Par.12	Continuity indicator					M									
Par.13	Event information						M								
Par.14	Forward call indicators								M						
Par.15	Information indicators									M					
Par.16	Information request indicators										M				
Par.17	Nature of connection indicators								M						
Par.18	Optional backward call indicators	O	O				O								
Par.19	Optional forward call indicators								O						
Par.20	Original called party number						O		O						
Par.21	Redirecting number								O						
Par.22	Redirection information								O			O			
Par.23	Redirection number											O			
Par.24	Suspend/resume indicators												M		M
Par.25	Transmission medium requirement								M						
Par.26	Transmission medium used	O	O		O		O								
Par.27	User service information								O						
Par.28	User-to-user information	O	O		O	O			O			O			

[a]M, Mandatory parameter; O, optional parameter.

Source: Rec. Q.763. Courtesy of ITU-T.

TABLE 11.2-3 Parameter Name Codes

Reference Number	Parameter Name	Name Code
Par.1	Access transport	0000 0011
Par.2	Application transport (ATP)	0111 1000
Par.3	Automatic congestion level	0010 0111
Par.4	Backward call indicators	0001 0001
Par.5	Call reference	0000 0001
Par.6	Called party number	0000 0100
Par.7	Calling party category	0000 1001
Par.8	Calling party number	0000 1010
Par.9	Cause indicators	0001 0010
Par.10	Circuit assignment map	0010 0101
Par.11	Closed user group interlock code	0001 1010
Par.12	Continuity indicator	0001 0000
Par.13	Event information	0010 0100
Par.14	Forward call indicators	0000 0111
Par.15	Information indicators	0000 1111
Par.16	Information request indicators	0000 1110
Par.17	Nature of connection indicators	0000 0110
Par.18	Optional backward call indicators	0010 1001
Par.19	Optional forward call indicators	0000 1000
Par.20	Original called party number	0010 1000
Par.21	Redirecting number	0000 1011
Par.22	Redirection information	0001 0011
Par.23	Redirection number	0000 1100
Par.24	Suspend/resume indicators	0010 0010
Par.25	Transmission medium requirements (TMR)	0000 0010
Par.26	Transmission medium used (TMU)	0011 0101
Par.27	User service information (USI)	0001 1101
Par.28	User-to-user information (UUI)	0010 0000

Note: Name codes are included only when the parameter is optional in the message.

Source: Rec. Q.763. Courtesy of ITU-T.

segmentation indicator fields allow segmentation and reassembly of data sequences that do not fit into a single message. SI is the *sequence indicator* (0 = subsequent segment, 1 = new sequence). The APM segmentation indicator identifies the segments in a sequence. SLR is the *segmentation local reference*.

Octets↓	8	7	6 5 4 3	2	1 ← Bits
1	ext	ACI			
2	ext	Spare		SNI	RCI
3	ext	SI	APM segmentation indicator		
3a	ext	SLR			
4a .. 4n	Encapsulated application information				

Figure 11.2-3. Format of Par.2: Application Transport (ATP). (From Rec. Q.763. Reproduced with the kind permission of ITU.)

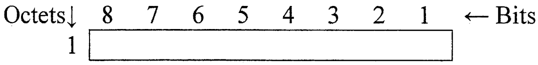


Figure 11.2-4. Format of parameters. Par.3: automatic congestion level. Par.7: calling party category. Par.13: event information. Par.25: transmission medium requirement.

Par.3 Automatic Congestion Level. This is an optional parameter in REL messages (Fig. 11.2-4). It indicates congestion at the exchange and is coded as follows:

0000 0001	Congestion level 1 exceeded (moderate congestion)
0000 0010	Congestion level 2 exceeded (severe congestion)

Par.4 Backward Call Indicators. This is a mandatory parameter in the ACM message and an optional parameter in CPG and ANM messages (Fig. 11.2-5). It contains a number of indicator bits with information for the originating exchange:

Bits		**Bits**	
BA	*Charge indicator*	*DC*	*Called party status*
00	No indication	00	No information
01	Do not charge	01	Subscriber free
10	Charge	11	Call setup delay
FE	*Called party category*	*HG*	*End-to-end signaling*
00	Not known	00	Not available
01	Subscriber	01	Pass-along method available
10	Payphone	10	SCCP method available
		11	Pass-along and SCCP methods available
J	*Interworking indicator*	*K*	*ISUP indicator*
0	TUP or ISUP signaling all the way (no interworking)	0	Not ISUP signaling all the way
		1	ISUP signaling all the way
1	Interworking encountered		
L	*Holding indicator*	*M*	*ISDN access indicator*
0	Holding not required	0	Called party has non-ISDN access (analog subscriber)
1	Holding required	1	Called party has ISDN access (ISDN user)
N	*Echo-control indicator*	*PO*	*SCCP method indicator*
0	Incoming half-echo control device not included	00	No indication
		01	Connectionless method available
1	Incoming half-echo control device included	10	Connection oriented method available
		11	Connectionless and connection-oriented methods available

Octets↓	8	7	6	5	4	3	2	1	← Bits
1	H	G	F	E	D	C	B	A	
2	P	O	N	M	L	K	J	I	

Figure 11.2-5. Format of parameters. Par.4: backward call indicators. Par.14: forward call indicators. Par.15: information indicators. Par.16: information request indicators. Par.22 redirection information.

Octets↓	8	7	6	5	4	3	2	1	← Bits
1	Call Identity								
2									
3									
4	Point Code								
5	Spare								

Figure 11.2-6. Format of Par.5: call reference. (From Rec. Q.763. Courtesy of ITU-T.)

The combination of bits H, G, J, and K is known as the *protocol control indicator*.

Par.5 Call Reference. This is an optional parameter in IAM, ACM, CPG, REL, SUS, RES, FOT, INF, and INR messages (Fig. 11.2-6). The parameter identifies a particular ISUP call at an exchange and consists of a *call identity* (assigned by the exchange that sends the message) and the point code of that exchange (two octets in ITU-T-ISUP; three octets in ANSI-ISUP). It is used in one form of end-to-end signaling (Section 11.5).

Par.6 Called Party Number. This is a mandatory variable-length parameter in IAM (Fig. 11.2-7). The parameter field consists of several subfields:

Odd/even indicator (*O/E*):

0 Even number of address digits
1 Odd number of address digits

Nature of address:

000 0001 Subscriber (directory) number
000 0011 National number
000 0100 International number

Numbering plan (only one value used):

001 ISDN/telephony numbering plan (see Section 1.2)

Bits H and D through A are not used.

Octets↓	8	7	6	5	4	3	2	1 ← Bits
1	O/E	Nature of Address						
2	H	Numbering Plan			D	C	B	A
3	2nd Digit				1st digit			
⋮								
n	Last even-numbered digit, or filler				Last odd-numbered digit			

Figure 11.2-7. Format of parameters. Par.6: called party number. Par.8: calling party number. Par.20: original called party number. Par.21: redirecting number. Par.23: redirection number. (From Rec. Q.763. Courtesy of ITU-T.)

Digits:

0000 to 1001	(Digit values 0–9)
1011	(Code 11)
1100	(Code 12)
1111	(End of address, ST)

If the number of address signals is odd, bits 8, . . . , 5 of octet n are filler coded 0000.

Par.7 Calling Party Category. This is a mandatory fixed-length parameter in IAM (Fig. 11.2-4), coded as follows:

0000 0001	French-speaking operator
0000 0010	English-speaking operator
0000 0011	German-speaking operator
0000 0100	Russian-speaking operator
0000 0101	Spanish-speaking operator
0000 1010	Ordinary subscriber
0000 1100	Voiceband data call
0000 1101	Test call

Octets↓	8	7	6	5	4	3	2	1 ← Bits
1	1	0	0	0	Location			
2	1	Cause Value						

Figure 11.2-8. Format of Par.9: cause indicators. (From Rec. Q.763. Courtesy of ITU-T).

Par.8 Calling Party Number. This is an optional parameter in IAM messages (Fig. 11.2-7). The coding is as in Par.6 (called party number), except that bits D, . . . , A are used:

	Bits
BA	*Screening indicator* (*SI*)
01	Calling number provided by calling user and verified by the network (originating local exchange)
11	Calling number provided by network

	Bits
DC	*Presentation restriction indicator* (*RI*)
00	Calling party allows the presentation of his number
01	Calling party does not allow the presentation of his number

Par.9 Cause Indicator. This is mandatory (MF) in REL messages and optional in ACM and CPG messages (Fig. 11.2-8). The Location field indicates where the message originated; the Cause field indicates why a call cannot be set up, or why a connection has to be released. The coding is identical to the coding of octets 3 and 4 in the Cause information element (IE.6) of Q.931 (Section 10.3.5).

Par.10 Circuit Assignment Map. This is an optional IAM parameter (Fig. 11.2-9) used to manage random assignment of channels in a multirate connection (Section 11.6.12). The *map type* field shows the type of digital multiplex (T1 or E1). The bits in octets 2–5 represent the individual channels: when a channel is used (not used) its bit is set to "1" ("0").

Octets↓	8	7	6	5	4	3	2	1 ← Bits
1	Spare		Map Type					
2	8	7	6	5	4	3	2	1
3	16	15	14	13	12	11	10	9
4	24	23	22	21	20	19	18	17
5	Spare	31	30	29	28	27	26	25

Map Type: 000001 = 1544kb/s 000010 = 2048 kb/s

Figure 11.2-9. Format for Par.10: circuit assignment map parameter field. (From Rec. Q.763. Reproduced with the kind permission of ITU.)

Octets↓	8	7	6	5	4	3	2	1 ← Bits
1	2nd Digit				1st Digit			
2	4th Digit				3rd Digit			
3	Binary Coded							
4	Number							

Figure 11.2-10. Format of Par.11: closed user group interlock code. (From Rec. Q.763. Courtesy of ITU-T.)

Par.11 Closed User Group Interlock Code. This is an optional IAM parameter (Fig. 11.2-10) used in countries that offer *closed user group* (CUG) service (Section 9.2.2 and 9.4.1). It uniquely identifies a CUG. Octets 1 and 2 contain four digits that identify the telecom that administers the interlock code. The binary number in octets 3 and 4 represents a particular code assigned by that telecom.

Par.12 Continuity Indicator. This is a mandatory (MF) parameter in COT messages (Fig. 11.2-11). Only indicator bit A is used:

Bit A	
0	Continuity check has failed
1	Continuity check successful

Par.13 Event Information. This is a mandatory (MF) parameter in CPG messages (Fig. 11.2-4). It indicates an event that has occurred during call setup.

0000 0001	Called party has been alerted
0000 0010	Progress
0000 0011	In-band tone or announcement is now available
0000 0100	Call forwarded on busy
0000 0101	Call forwarded on no answer
0000 0110	Call forwarded unconditionally

Octets↓	8	7	6	5	4	3	2	1 ← Bits
1	H	G	F	E	D	C	B	A

Figure 11.2-11. Format of parameters. Par.12: continuity indicators. Par.17: nature of connection indicators. Par.18: optional backward call indicators. Par.19: optional forward call indicators. Par.24: suspend/resume indicators.

Par.14 Forward Call Indicators. This is a mandatory (MF) parameter in IAM messages (Fig. 11.2-5). It contains indicator bits with information for the intermediate and terminating exchanges in a connection.

Bits		Bits	
A	*Call type*	*CB*	*End-to-end signaling*
0	Call to be treated as national call	00	Not available
1	Call to be treated as international call	01	Pass-along method available
		10	SCCP method available
		11	Pass-along and SCCP methods available
D	*Interworking indicator*		
0	No interworking encountered (TUP or ISUP signaling all the way)		
1	Interworking encountered		
F	*ISUP indicator*	*HG*	*ISUP preference indicator*
0	ISUP signaling not used all the way	00	ISUP signaling preferred all the way
1	ISUP signaling used all the way	01	ISUP signaling not required all the way
		11	ISUP signaling required all the way
I	*ISDN access indicator*		
0	Calling party has non-ISDN access (analog subscriber)		
1	Calling party has ISDN access (ISDN user)		

The combination of bits C, B, D, F, H, and G constitutes the *protocol control indicator* (PCI).

Par.15 Information Indicators. This is a mandatory parameter in INF (Information) messages (Fig. 11.2-5). It consists of bits representing the parameters or call-control functions that can be requested by an exchange. The bit values indicate whether the corresponding parameter is included in the message, or whether the corresponding service is being provided. For example:

Parameter or Function	Bit(s)	Value(s)	Meaning
Calling party number (Par.8)	BA	00	Not included
		01	Not available
		11	Included
Call holding	C	0	Not provided
		1	Provided
Calling party category (Par.7)	F	0	Not included
		1	Included

Par.16 Information Request Indicators. This is a mandatory parameter in the INR (Information Request) message (Fig. 11.2-5). It consists of indicator bits that represent parameters or call-control functions. The bit value indicates whether the corresponding parameter or function is requested.

Parameter or Function	Bit(s)	Value(s)	Meaning
Calling party number (Par.8)	A	0	Not requested
		1	Requested
Call holding	B	0	Not requested
		1	Requested
Calling party category (Par.7)	E	0	Not requested
		1	Requested

Par.17 Nature of Connection Indicators. This is a mandatory (MF) parameter in IAM messages (Fig. 11.2-11).

Bits	
BA	*Satellite indicator*
00	No satellite circuit in connection
01	One satellite circuit in connection
10	Two satellite circuits in connection

Bits	
DC	*Continuity check indicator*
00	Continuity check not required
01	Continuity check required on this circuit

Bit	
E	*Echo-control device indicator*
0	No outgoing half-echo control device included
1	Outgoing half-echo control device included

Bits H, G, and F are not used.

Par.18 Optional Backward Call Indicators. This is an optional parameter in ACM, CPG, and ANM messages (Fig. 11.2-11). Only bit A is used.

Bit	
A	*In-band information indicator*
0	No indication
1	In-band tone or announcement is now available

Par.19 Optional Forward Call Indicators. This is an optional parameter in IAM messages, used in countries that offer CUG service (Fig. 11.2-11). Only bits B and A contain information:

Bits BA	
00	Non-CUG call
10	CUG call; outgoing access allowed
11	CUG call; outgoing access not allowed

Par.20 Original Called Party Number. This optional parameter is included in IAM messages of calls that have been redirected to a new called number. It has the format of Par.8 (calling party number)—see Fig. 11.2-7. The presentation screening indicator (bits D, C) pertains to the original called number. If no presentation is allowed, octets 3 to *n* are omitted. The screening indicator (bits B, A) is not used.

Par.21 Redirecting Number. This optional parameter is included in IAM messages of calls that are being forwarded. It is the number of the party that caused the call to be forwarded and has the format of Par.20.

Par.22 Redirection Information. This is an optional parameter, included in IAM and REL messages on calls that are being forwarded or rerouted—see Fig. 11.2-5.

Bits C, B, and A indicate the type of redirection, for example:

001	Call rerouted
010	Call forwarded

Bits H, G, F, and E indicate the original redirection reason:

0000	Unknown
0001	User busy
0010	No answer
0000	Unconditional call forwarding

Bits K, J, and I hold a binary number between 1 and 5, which indicates the number of redirections the call has undergone.

Par.23 Redirection Number. Suppose that an exchange has received the IAM of an incoming call and determines that the connection has to be rerouted. It then sends a (backward) REL message, which may include a redirection number, which is the new called number. It has the format and coding of Par.6 (called party number).

Par.24 Suspend/Resume Indicators. This is a mandatory parameter in SUS (Suspend) and RES (Resume) messages. It indicates the originator of the message —see Fig. 11.2-11:

Bit A	Originator
0	ISDN user
1	Network (exchange)

Par.25 Transmission Medium Requirement (TMR). This is a mandatory IAM parameter and indicates the nature of the information in the call. It is used by exchanges to select outgoing trunks of the appropriate type—see Fig. 11.2-4:

0000 0000	Speech
0000 0010	64-kb/s Data unrestricted
0000 0011	3.1-kHz Audio
0000 0110	64-kbits/s Preferred
0000 0111	2 × 64-kb/s Unrestricted
0000 1000	384-kb/s Unrestricted
0000 1001	1564-kb/s Unrestricted
0000 1010	1920-bits/s Unrestricted
0001 0000-1111	$N \times$ 64-kb/s Unrestricted ($N = 3$ to 18)*
0010 0000-1010	N × 64-kb/s Unrestricted ($N = 19$ to 29)*

* 0 0 0 1 0 0 1 (6 × 24) and 0 0 1 0 0 1 0 1 (26 × 24) are not used (spare).

When the *fallback* feature is used (Section 11.6.11), a second instance of this parameter, TMR′, is used to convey the fallback value. TMR′ values range only from speech to 1920-kb/s unrestricted.

Par.26 Transmission Medium Used (TMU). This is an optional parameter in ACM, ANM, CPG, and CON messages. It indicates that the fallback procedure has taken place (Section 11.6.11). The format and values are the same as for TMR.

Par.27 User Service Information (USI). This is an optional IAM parameter, included only in calls originated by an ISDN user. It is coded as octets 3–5a of the bearer capability information element (IE.1) of Q.931 (Section 10.3.5). When the fallback feature is used (Section 11.6.11), a second instance of this parameter, USI′, is used to convey the fallback value.

Par.28 User-to-User Information (UUI). This is a variable-length optional IAM parameter, used in countries that offer user-to-user signaling to ISDN users. The parameter contents are the same as in Rec. Q.931 (Chapter 10).

11.3 SIGNALING FOR CALLS BETWEEN ISDN USERS

This section describes the Q.931 and ISUP signaling for a typical successful inter-exchange call between two ISDN users, with ISUP signaling all the way [6].

11.3.1 Connection Setup

We first examine the setup of the connection (Fig. 11.3-1). Trunks T_1 and T_2 have ISUP signaling. As in TUP signaling, an exchange decides whether to do a continuity check on the seized outgoing trunk, and informs the incoming exchange (Section 9.3.1). In this example, no checks are made.

The focus is on the action at the exchanges and the interworking between Q.931 and ISUP signaling messages. We refer to the ISUP and Q.931 messages by their acronyms (Sections 11.2.3 and 10.3.2).

User U_1 originates the call at TE_1, with a SETUP message that includes the complete called address (*en-bloc* address signaling). Exchange P examines the message, seizes the outgoing trunk T_1 to exchange Q, and sends an IAM. It also assigns a B-channel to TE_1, cuts through a unidirectional switchblock path, and sends a CALPRC message to TE_1.

After receiving the IAM, exchange Q seizes outgoing trunk T_2 to exchange R, sends an IAM, and cuts through a bidirectional path between T_1 and T_2.

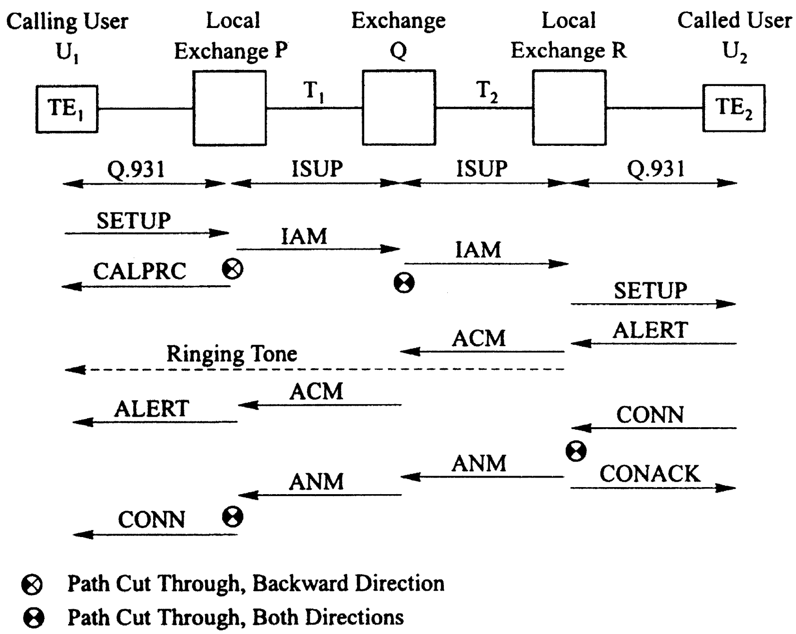


Figure 11.3-1. Set-up of a call between ISDN users.

When exchange R receives the IAM, it assigns a B-channel on the DSL of U_2 and broadcasts a SETUP message to the TEs on the DSL. In this example, TE_2 responds with an ALERT message and alerts user U_2 with an audible or visual signal. Exchange R then sends an ACM message to exchange Q. If Par.27 (user service information) indicates a speech or 3.2-kHz audio call, exchange R also connects trunk T_2 to a ringing-tone source, and calling user U_1 then hears the tone. Exchange Q repeats the ACM to exchange P, which then sends an ALERT message the TE_1.

When U_2 answers, TE_2 sends a CONN message. Exchange R then cuts through a bidirectional path between T_2 and the B-channel allocated to TE_2, sends a CONACK (acknowledgment) to TE_2, and an ANM message to exchange Q. The message is repeated backwards and reaches exchange P. That exchange then cuts through its forward transmission path and sends a CONN message to TE_1. The TE may or may not acknowledge the CONN message, depending on its type and manufacturer.

The conversation or data transfer between U_1 and U_2 can now begin.

11.3.2 Release of the Connection

At the end of the call, the B-channels used by TE_1 and TE_2 and trunks T_1 and T_2 have to be released. In Q.931 and ISUP signaling, either user can initiate the release. In this example, U_2 initiates the release and commands TE_2 to send a DISC message (Fig. 11.3-2). Exchange R then releases the B-channel to TE_2 at its end and returns a RLSE message. TE_2 then releases the B-channel at its end and sends a RLCOM message. This completes the release and the B-channel is now available for new calls.

Exchange R also sends a backward REL message for trunk T_2 to exchange Q and starts to clear the trunk at its end. In turn, exchange Q sends a REL message for trunk

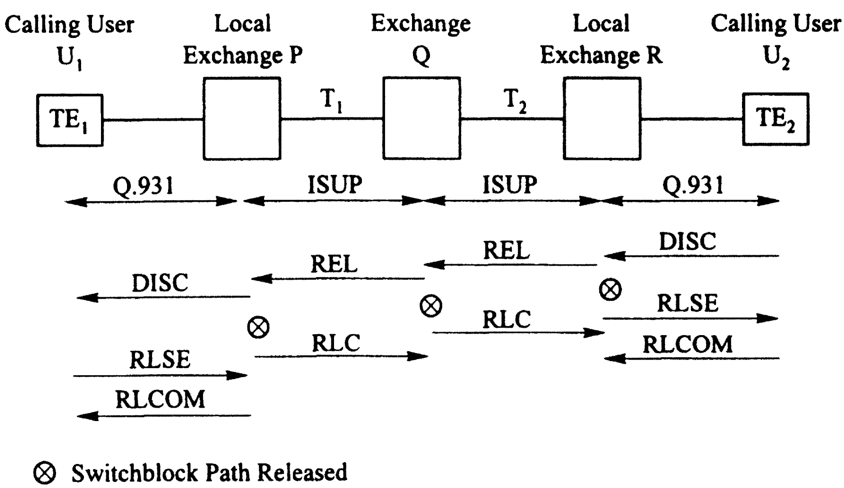


Figure 11.3-2. Release of the connection.

T_1 to exchange P and starts to clear the trunk. After releasing, respectively, T_2 and T_1 at their end, exchanges Q and P return RLC messages, indicating to exchanges R and Q that trunks T_2 and T_1 are available again for new calls.

In ISUP signaling, an exchange first sends a REL message for a trunk and then clears the trunk. This results in a faster release than in TUP signaling, where an exchange first clears a trunk and then sends a clear-forward signal (Section 9.3.1).

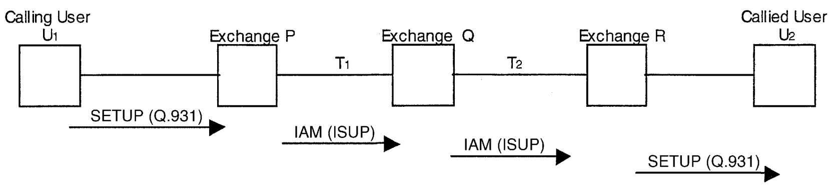
When exchange P receives the REL message, it also requests TE_1 to release its B-channel with a DISC message. TE_1 then clears the channel at its end and returns a RLSE message. In response, exchange P releases the B-channel at its end and sends a RLCOM message. The B-channel is now available for new calls.

11.3.3 Q.931 $\leftrightarrow$ ISUP Interworking and Mapping

In the example of Figs. 11.3-1 and 11.3-2, interworking between Q.931 signaling and ISUP signaling takes place in local exchanges P and R. These exchanges have to "map" received Q.931 messages and information elements (IEs) to outgoing ISUP messages and parameters (Par.)—and vice versa. The IEs and parameters have been described in Sections 10.3.5 and 11.2.4. We now examine the mappings for the call [2].

Mapping of SETUP and IAM Messages (Fig. 11.3-3). Exchange P receives information elements IE.1 (mandatory) and IE.2, IE.3, IE.4, IE.5, IE.8, IE.10 (optional) from calling ISDN user U_1.

The complete information in bearer capability (IE.1) is mapped into Par.27 (user service information), which is transferred transparently across the network. A part of



IE.1 Bearer Capability (M)	Par.25 Transmission Medium Requirements (M)	—
	Par.27 User Service Information (O)	IE.1 (M)
IE.2 Called Party Number (M)	Par.6 Called Party Number (M)	IE.2 (M)
IE.3 Calling Party Number (O)	Par.8 Calling Party Number (O)	IE.3 (O)
IE.4 Called Party Subaddress (O)	Par.1 Access Transport (O)	IE.4 (O)
IE.5 Calling Party Subaddress (O)		IE.5 (O)
IE.8 High Layer Compatibility (O)		IE.8 (O)
IE.10 Low Layer Compatibility (O)		IE.10 (O)
—	Par.7 Calling Party Category (M)	—
—	Par.17 Nature of Connection Indicators (M)	—
—	Par.14 Forward Call Indicators (M)	IE.11 Progress Indicator (O)

(M): mandatory parameter or IE (O): optional parameter or IE

Figure 11.3-3. Mappings of Q.931 SETUP messages and ISUP IAM message.

the information in IE.1 is mapped into Par.25 (transmission medium requirements), which is used by the exchanges during the setup of the connection.

IE.2 and IE.3 (called and calling party address) are mapped into Par.6 and Par.8. IE.2 is shown as optional because U_1 may not include it in the SETUP messages and may send the digits of the called party number in a series of INFO messages (Section 10.5.3). Our example assumes that U_1 has included IE.2 in the SETUP message. IE.3 is also optional. In countries where Par.8 is mandatory, if exchange P has not received IE.2, it enters the default directory number of U_1, which is a stored parameter at the exchange, into Par.8.

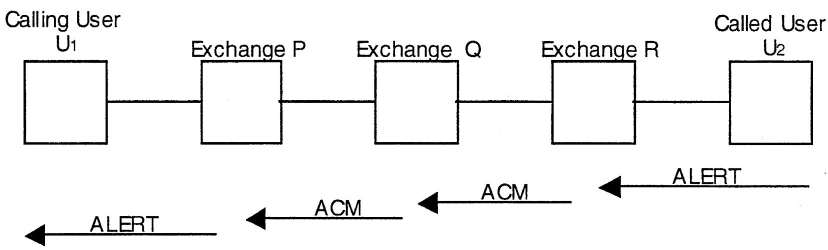
If included in the SETUP message, IE.4, IE.5, IE.8, and IE.10 are entered into Par.1 (access transport). This parameter is transferred transparently across the network.

The contents of parameters Par.7, Par.17, and Par.14 are set by exchange P.

We next explore the mapping—at terminating exchange R—of the IAM parameters into information elements that have to be included in the SETUP message to U_2. Par.27 is always mapped into IE.1. Par.6 (called party number) is mapped into IE.3 if U_2 has "calling line identity presentation service." The information elements present in Par.1 are retrieved and included in the SETUP message.

Under conditions to be discussed later, some indicators of Par.14 (forward call indicators) are mapped to IE.11 (progress indicators) in the SETUP message to U_2.

Mapping of ALERT and ACM Messages (Fig. 11.3-4). The ALERT message received from user U_2 normally does not include IEs. The corresponding ACM messages include a Par.4 (backward call indicators), whose values are set by exchange R. Par.4 is processed by exchanges Q and P, and under certain conditions, an indicator is mapped into IE.11 (progress indicators), which is included in the ALERT message to user U_1.



IE 11 Progress Indication (O)	Par.4 Backward Call Indicators (M)	------

(M): mandatory parameter (O): Optional IE

Figure 11.3-4. Mapping of Q.931 ALERT messages and ISUP Address Complete message (ACM).

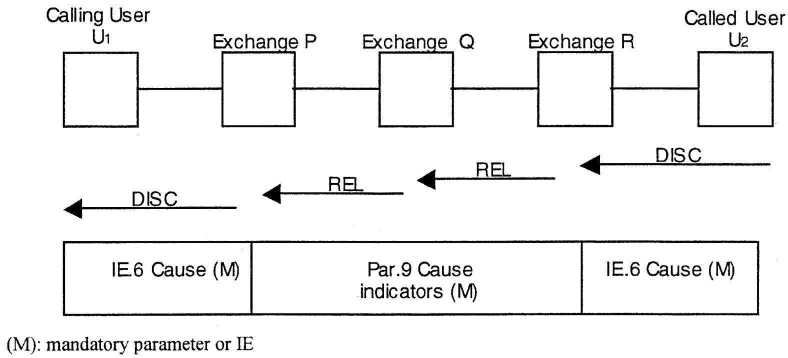


Figure 11.3-5. Mapping of Q.931 DISC messages and ISUP Release message (REL).

Mapping of DISC and REL Messages (Fig. 11.3-5). DISC and REL messages are general-purpose requests to release a B-channel or a trunk. They always include, respectively, an IE.6 (cause indicators) and Par.9 (cause indicators), to indicate the reason for the release. Mappings between IE.6 and Par.9 take place at exchanges P and R.

The coding of the value fields of the information elements and parameters can be found in Sections 10.3.4 and 11.2.4. Some mappings require format changes. For example, the digits of the called and calling numbers have a seven-bit format in IE.2 and IE.3, and a four-bit format in Par.6 and Par.8. Other mappings, for example, between IE.6 (cause) and Par.9 (cause indicators), do not require reformatting.

11.3.4 Use of Message Parameters

This section revisits the call example described in Sections 11.3.1 and 11.3.2, focusing this time on the ISUP message parameters and their use. It is helpful to refer to the parameter descriptions of Section 11.2.4 when reading this material.

Routing Parameters in IAM. The selection of outgoing trunk groups by exchanges P and Q is influenced by Par.6, Par.25, Par.17, and Par.14 of the IAM (Fig. 11.3-3).

Par.6 (called party number) is the primary routing parameter, from which the exchange derives the route set for the call. The other parameters can disqualify some outgoing trunk groups in the route set.

If Par.25 (transmission medium requirements) indicates that the transmission medium should accommodate a 64-kb/s data call, the connection is restricted to exchanges with digital switchblocks and to digital trunk groups without circuit multiplication equipment.

Par.17 (nature of connection indicators) is used to limit the number of satellite trunks in the connection, to control the inclusion (enabling) of echo-control devices, and to indicate whether a continuity check will be made on the trunk.

In Par.14 (forward call indicators), bits CB, D, F, and HG constitute the protocol control indicator (PCI). Bits HG (ISUP preference indicators) specify whether ISUP signaling is required all the way. Originating exchanges are usually programmed to indicate that all-the-way ISUP signaling is required for calls originated by ISDN users and preferred for calls originated by subscribers. Indicators HG are not modified by later exchanges along the connection. Indicators D (interworking indicator) and F (ISUP indicator) indicate the actual signaling on the connection. If exchange P in Fig. 11.3-1 has set HG, D, and F to "ISUP required all the way," "no interworking encountered," and "ISUP signaling used all the way," exchange Q selects an outgoing trunk with ISUP signaling and does not change these indicators in its IAM to exchange R. The settings of these indicators in connections where interworking occurs between ISUP and other interexchange signaling systems are discussed in Section 11.4. The meaning of indicators CB is discussed in Section 11.5.

Call Processing Information at the Originating and Terminating Exchange. In the call of Fig. 11.3-1, the local exchanges know that the signaling is ISUP all the way. Terminating exchange R receives this information in bit F of Par.14 in the IAM, and originating exchange P is informed by bit K of Par.4 (backward call indicators) of the ACM message. We shall see later that, if the signaling is not ISUP all the way, both local exchanges are also made aware that this is the case. This information is sometimes needed for subsequent call-processing actions.

In connections with ISUP signaling all the way, both local exchanges also know whether the distant party in the connection is an ISDN user or a subscriber from, respectively, bit I of Par.14 (in IAM) and bit M of Par.4 (in ACM). In the call of Fig. 11.3-1, both parties are ISDN users. In calls between an ISDN user and a subscriber, the ISDN user is informed by his local exchange that the distant party is non-ISDN. This is done by including an IE.11 (progress indicator) in a Q.931 message to the ISDN user (Section 11.4).

Information Included in DISC and REL Messages. The DISC and REL messages always include, respectively, an IE.6 (cause) and Par.9 (cause indicator). In the example of Fig. 11.3-2, the location and cause values indicate "public network of the remote user" and "normal call clearing" (Section 10.3.5).

11.4 CALLS INVOLVING ANALOG SUBSCRIBERS

Presently, the number of ISDN users is small, and the majority of calls carried on ISUP trunks are between two analog subscribers. This section explores ISUP signaling for calls in which one or both parties are subscribers.

11.4.1 Call Between Two Subscribers

Figure 11.4-1 shows a call between subscribers in which trunks T_1 and T_2 have ISUP signaling.

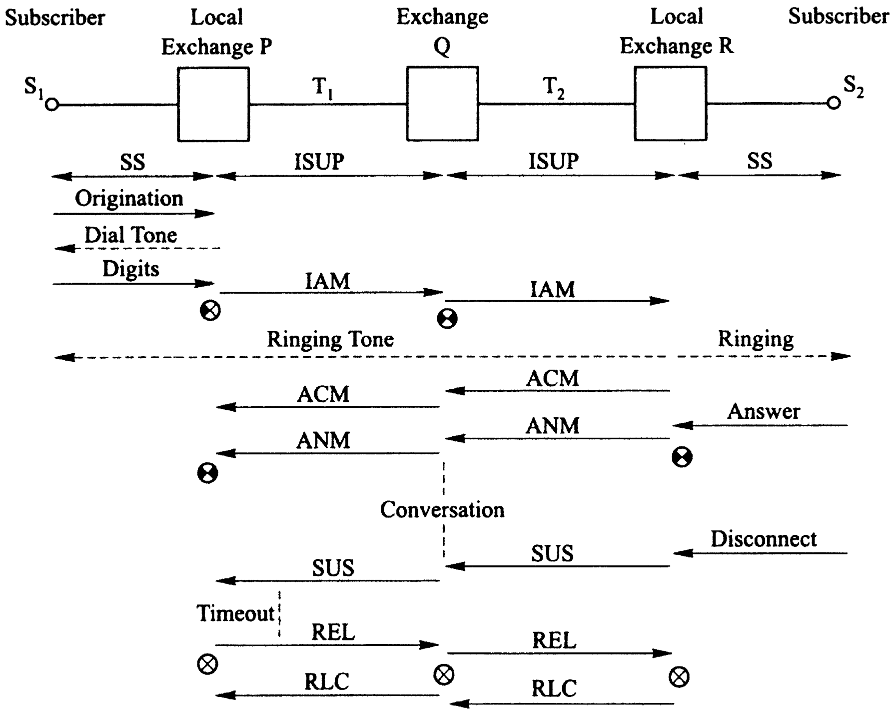


Figure 11.4-1. Call between subscribers. ISUP signaling all the way.

Setup of Connection. Subscriber S_1 has originated a call. We first examine the IAM sent by the originating exchange P (Table 11.2-2). Par.6 holds the called party number dialed by S_1. Par.1 (access transport) and Par.27 (user service information) are absent because subscriber signaling does not provide bearer capability, subaddresses, and compatibility information.

The originating exchange cannot determine whether the call will be speech or 3.1-kHz audio. Local exchanges are usually programmed to set Par.25 (transmission medium requirement) to "3.1-kHz audio."

The calling party category (stored in memory at the exchange) is entered into Par.7. We assume here that the local exchanges are programmed to include Par.8 (calling party number).

In Par.14 (forward call indicators), bits D, F, and I are set to "no interworking encountered," "ISUP signaling used all the way," and "calling party has non-ISDN access." In calls originated by subscribers, bits HG are usually set to "ISUP signaling preferred all the way." In this example, exchange Q selects outgoing trunk T_2, which has ISUP signaling, and does not change the values of Par.14 in its IAM to exchange R.

When a call setup arrives at the terminating exchange on an ISUP trunk, the exchange always checks whether the called party is compatible with the service indicated in Par.25 (transmission medium requirement). In this example the called party is a subscriber and is compatible with the "3.1-kHz audio service." Exchange R therefore proceeds with the call setup.

Assuming that S_2 is idle, it rings the subscriber line, connects trunk T_2 to a ringing-tone source, and sends an ACM message to exchange Q, which repeats it to exchange P.

In the ACM, bits DC, FE, and M of Par.4 (backward call indicators) are set to "subscriber free," "subscriber," and "called party has non-ISDN access." Bits J and K are copied from bits D and F in Par.14 (forward call indicators) of the received IAM and indicate "no interworking encountered" and "ISUP signaling all the way." Exchange P is therefore aware that signaling is ISUP all the way and that the called party is non-ISDN (it may need this information for later call-control actions).

Since exchange R receives the calling party number (Par.8) in the IAM, it can provide CLASS services to the called subscriber (Section 3.5.2).

When exchange R receives an answer signal from S_2, it cuts through the path between T_2 and the called subscriber line and sends an ANM message. When the message reaches exchange P, it cuts through its forward transmission path and the conversation or data transmission can start.

Release of the Connection. Figure 11.4-1 assumes that the called subscriber disconnects first. Called subscribers have a grace period that allows them to disconnect the telephone on which they answered the call, and to pick up another telephone, without causing the release of the connection (Section 3.1.1). Therefore, when S_2 disconnects, exchange R sends a SUS message (a REL message would initiate the release of the network connection). On receipt of the message, exchange P starts a timer with a timeout of, say, 30 s. If S_2 re-answers before the timer expires, exchange R sends a RES (Resume) message, exchange P stops the timer, and the call continues. In the example, exchange P has not received a RES message when the timer expires and initiates the release of the connection.

11.4.2 Calls Between a User and a Subscriber

We now consider calls between a user and a subscriber. The call configuration is as in Fig. 11.4-1, except that one party is a subscriber and the other party is an ISDN user. Since the connection has ISUP signaling all the way, terminating exchange R always knows whether the remote (calling) party is a subscriber or an ISDN user from bit I of Par.14 (forward call indicators) in the IAM message. Likewise, originating exchange P knows the type of the remote (called) party from bit M of Par.4 (backward call indicators) in the ACM message.

Since some user actions depend on whether the distant party is a user or a subscriber, a local exchange always informs its user if the distant party is non-ISDN. This is done by including an IE.11 (progress indicator) in the CONN message to the TE of a calling user, or in the SETUP message to the DSL of the called user. The TE then presents this information to its user.

In calls from a subscriber to a user, the IAMs do not include a Par.1 (access transport), which contains IEs for called TE selection. If the called user has terminals of several types, the only method for the selection of a called TE is by multiple directory numbers (Section 10.4.3).

11.4.3 Connections with Signaling Interworking

Figure 11.4-2 shows two connections in which one trunk has ISUP signaling and the other trunk has multifrequency (MF) signaling (Chapter 4). Connections involving MF-ISUP interworking can be used for calls originated by a subscriber and for calls originated by a user who has requested a speech or 3.1-kHz audio call and has not included in the SETUP message information elements that require ISUP signaling.

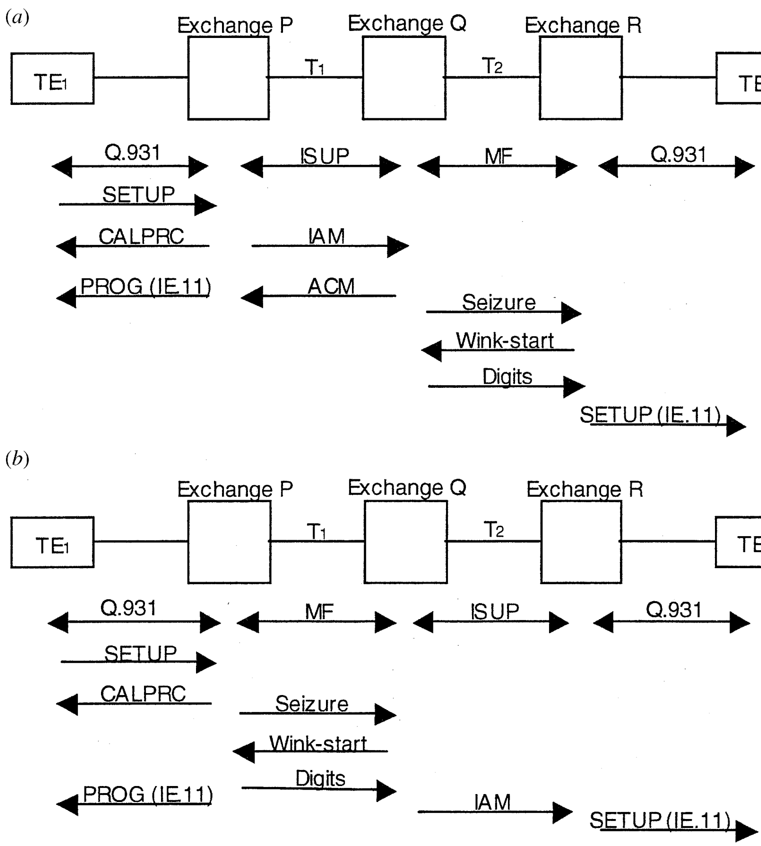


Figure 11.4-2. Connection with ISUP-MF interworking.

In these calls, a calling or called TE is always notified by its local exchange that the signaling on the connection is not ISUP all the way, and that therefore no information is available regarding the type (subscriber or user) of the distant party. The notification is in a message with an IE.11 (progress) set to "not ISUP signaling all the way."

In example (a), exchange P has indicated—in bits HG of Par.14 (forward call indicators)—that ISUP signaling all the way is preferred, or not required, and exchange Q has selected an outgoing MF trunk T_2. After seizing the trunk, Q returns an ACM for trunk T_1 in which bit K of Par.4 (backward call indicators) is set to "not ISUP signaling all the way." Exchange P then notifies TE_1 with a PROG message that includes IE.11. Also, since the called party is an ISDN user and the call has arrived at exchange R on a MF trunk, the exchange includes IE.11 in its SETUP broadcast message on the called DSL.

In example (b), exchange P has selected an outgoing MF trunk and informs TE_1 with a PROG message that includes IE.11. Moreover, exchange Q sets bit F for Par.14 (forward call indicators) in its IAM to "not ISUP signaling all the way," and exchange R then includes the IE.11 in its SETUP message.

11.5 END-TO-END SIGNALING

ISUP signaling includes procedures for end-to-end signaling between the originating and terminating exchanges of a connection. The applications of ISUP end-to-end signaling are usually related to supplementary services and are similar to those of MFC-R.2 and TUP (Chapters 4 and 9). ISUP end-to-end signaling is possible only on connections in which all trunks have ISUP signaling. In this section we explore this signaling in a connection from originating exchange A to terminating exchange C, and routed via intermediate exchange B.

ISUP has two methods for end-to-end signaling: the pass-along method and the SCCP (Signaling Connection Control Part) method [2, 6]. In some networks only one of these methods is implemented; in other networks, both methods exist side by side.

11.5.1 Information Request and Information Messages

In ISUP end-to-end signaling, the exchange at one end of the connection requests information, or an action, with an Information Request (INR) message. The exchange at the other end responds with an Information (INF) message.

The INR messages include a mandatory Par.16 (information request indicators) whose bits represent individual information items and actions. The bits indicate that the corresponding information or action is being requested (1), or not requested (0). For example, the terminating local exchange can request the originating local exchange to hold the connection until the terminating exchange sends a DISC message.

The INF messages include a mandatory Par.15 (information indicators) whose bits show whether a requested information item or action is provided. When an information item is provided, it is included as an optional parameter in the message.

11.5.2 Call Indicators

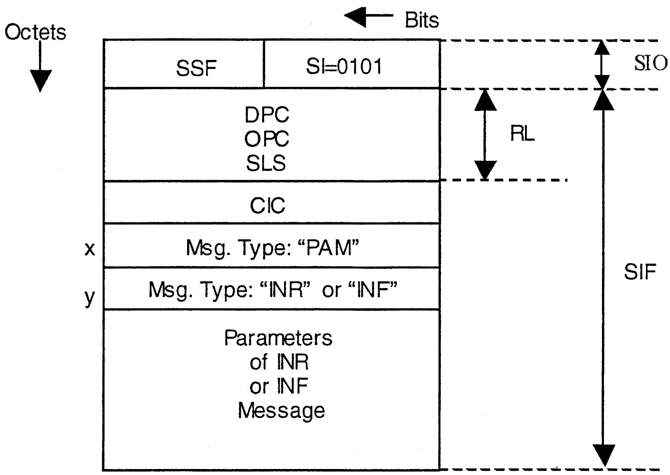
Early in the call setup, the calling and called exchanges inform each other about their end-to-end signaling capabilities. The originating local exchange indicates its capabilities with bits C and B of Par.14 (forward call indicators) in its IAM, and the terminating local exchange does the same with bits H and G of Par.4 (backward call indicators) in its first backward message (CPG, ACM, or ANM) for the call.

11.5.3 Pass-Along Method

In this method the originating and terminating local exchanges place their outgoing INF and INR messages in a Pass-Along message (PAM), shown in Fig. 11.5-1. Octet x identifies the message as a PAM, and octet y indicates the type of embedded message.

Pass-Along messages are transferred—forward or backward—by the exchanges along the connection for the call, just like other ISUP messages. Figure 11.5-2 shows the transfer of a PAM from originating exchange A to terminating exchange C, for a call routed on trunks T_1 and T_2.

The point codes of exchanges A, B, and C are a, b, and c. The circuit identification codes of the trunks are cic_1 and cic_2. The figure shows only the parameters DPC (destination point code), OCP (originating point code), and CIC (circuit identification code), in the MTP-transfer primitives.



SIO: service information octet. SIF: signaling information field.
RL: routing label. CIC: circuit identification code

Figure 11.5-1. Pass-along message.

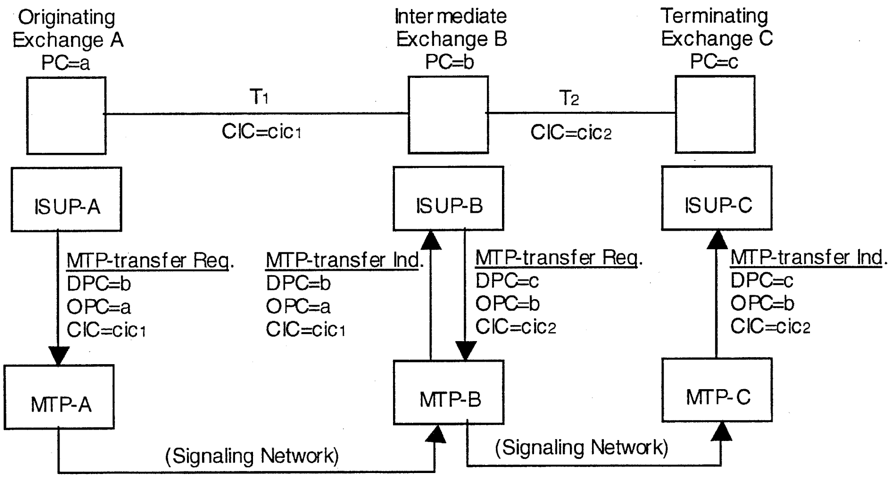


Figure 11.5-2. End-to-end signaling, pass-along method.

ISUP-A passes the PAM to MTP-A, in an MTP-transfer request that includes the parameters DPC = b, OPC = a, and CIC = cic_1, which identify trunk T_1. The MTP at exchange B passes the received message to ISUP-B. A call record at the exchange shows that T_1 is connected to T_2. ISUP-B does not examine the contents of the embedded INF or INR message and returns the PAM to MTP-B with parameters DPC = c, OPC = c, and CIC = cic_2, which identify trunk T_2. MTP-C passes the message to ISUP-C, which then knows that the message relates to the call on trunk T_2.

11.5.4 End-to-End Signaling, SCCP Method

In the SCCP method, ISUP uses the services of SCCP, which is a part of Signaling System No.7 and is described in Chapter 15. At this point we only need to know that messages between ISUP and SCCP are passed in N-unitdata primitives, which include the point code of the message destination.

Figure 11.5-3 shows the transfer of a message from originating exchange A to terminating exchange C. ISUP-A passes its outgoing INF or INR message to SCCP-A in an N-unitdata request that includes a subsystem number (SSN = "ISUP"), which identifies the message as an ISUP message, the point code of the destination exchange (DPC = c), and the INF or INR message. SCCP-A embeds the message in a SCCP message and passes it to MTP-A, in an MTP-transfer request primitive that includes the message and the point code of exchange C.

MTP-C passes the received SCCP message to SCCP-C, which extracts the embedded INF or INR message. Because SSN = "ISUP," SCCP passes the message to ISUP-C, with an N-unitdata indication.

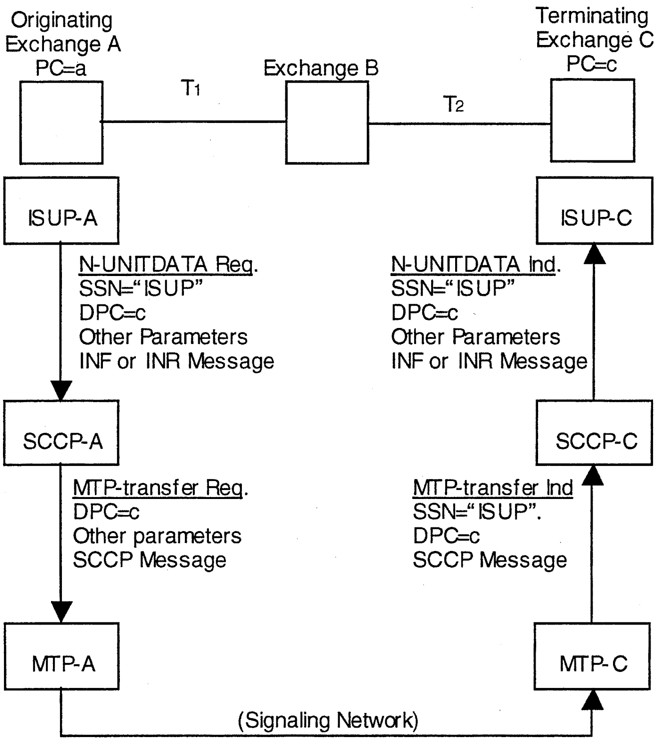


Figure 11.5-3. End-to-end signaling, SCCP method.

Call References. In SCCP end-to-end signaling, exchanges A and C have to know each other's point codes and need a mechanism to identify the call to which a INF or INR message pertains. This is done by including a Par.5 (call reference) in the message. The parameter consists of a point code (PC) and a call identity (CID).

When exchange A sends its IAM for the call, it assigns a call identity (CID = x) and includes a Par.5 with PC = a and the assigned CID (Fig. 11.5-4).

When exchange C receives the IAM, it associates A's call reference with the call. It also assigns a call identity (CID = y) and, in its first backward message, includes a Par.5 with PC = c and the assigned CID. On receipt of the message, exchange A associates C's call reference with the call.

Exchanges A and C now know each other's call references. Then, if exchange A needs to send an INR message, it includes a Par.5 with C's call reference and passes the message to its SCCP with an N-unitdata request, identifying the message destination by DPC = c. Likewise, exchange C responds with an INF message that includes a Par.5 with A's call reference and passes it to its SCCP, identifying the message destination by DPC = a.

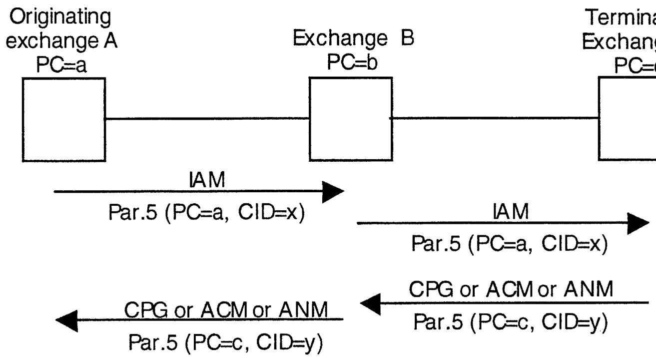


Figure 11.5-4. Establishing call references.

11.6 OTHER SIGNALING PROCEDURES

This section outlines several ISUP signaling procedures that have not yet been discussed. Some of these are the counterparts of the TUP procedures described in Sections 9.3 and 9.5.

11.6.1 Continuity Checking

As in TUP signaling, an exchange performs a continuity check on a seized outgoing ISUP trunk under certain conditions. When an exchange, say, exchange A, seizes an outgoing trunk to exchange B, indicators in Par.17 (nature of connections indicators) of its IAM specify whether the continuity check will be made. If so, exchange B establishes a loop-back on the incoming trunk.

When exchange A has made a continuity check, it sends a COT (Continuity) message to exchange B. In the message, Par.12 (continuity indicator) indicates whether the check has been successful.

11.6.2 Calling Line Identity Presentation

This service, in which the calling party number is presented to the called party, can be provided in two ways.

A telecom may program the exchanges in its network to always include Par.8 (calling party number) in their IAMs.

Alternatively, the local exchanges in a network can be programmed to obtain the calling party number on calls that terminate to subscribers who have the service. The terminating exchange then sends an Information Request (INR) message to the originating exchange, using end-to-end signaling. The message contains a Par.16 (information request indicators), which is set to request the calling party number. The originating exchange returns an Information (INF) message that includes a Par.15 (information indicators) and a Par.8 (calling party number).

11.6.3 Closed User Group Service

In this service, the local exchange of the called subscriber allows or disallows the incoming call, based on closed user group (CUG) information about the calling and called parties (Section 9.4.7). The CUG information of the calling party is transferred by the IAM. Par.11 (closed user group interlock code) specifies the caller's CUG, and Par.19 (optional forward call indicators) indicates whether the calling party is allowed to access a party outside his CUG. The CUG information about the called party is available at the terminating local exchange.

11.6.4 User-To-User Signaling

In countries that offer this service, an ISDN user can exchange user-to-user information during the setup and release of a connection, by including a user–user IE in a Q.931 SETUP, ALERT, CONN, and/or DISC message (Section 10.3.5).

A local exchange maps the IE.14 received in a Q.931 message into Par.28 (user-to-user information, UUI) and includes the parameter in a sent ISUP message (Fig. 11.6-1). The other local exchange maps a received Par.28 into an IE.14 and includes it in a Q.931 message.

11.6.5 Call Forwarding

Figure 11.6-2 is an example of call forwarding in a network with ISUP trunks. Customer C_1 (an analog subscriber or ISDN user) initiates a call to customer C_2, who is currently forwarding incoming calls to customer C_3. In turn, this customer is forwarding incoming calls to C_4. The numbers of customers C_2, C_3, and C_4 are N_2, N_3, and N_4.

We examine the parameters relating to call forwarding in the IAMs sent by exchanges A, B, and C. The called party number (Par.6) is the number as perceived by the exchange sending the IAM. The original called party number (Par.20) is the number dialed by C_1. The redirecting number (Par.21) is the number of the redirecting (forwarding) party. Finally, the number of redirections (an item in Par.22) shows how often the call has been redirected.

This information is used by the exchange that receives a forwarded call to decide whether to forward the call again. In the first place, the call should not be forwarded to the original called number or to the redirecting number. Also, the exchange may be programmed to not forward the call if it already has undergone a predetermined maximum number of redirections.

The forwarding exchange also sends a backward Call Progress (CPG) message, which is repeated to the originating exchange A. The message includes a Par.13 (event information), which indicates the reason why the call has been forwarded, and a Par.21 (redirecting number), which represents the customer who is requesting

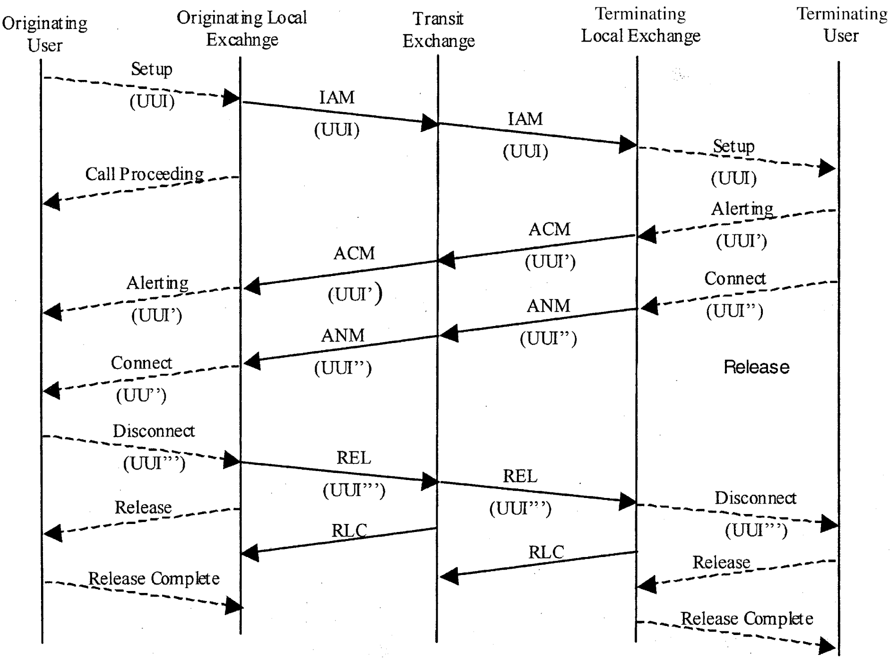
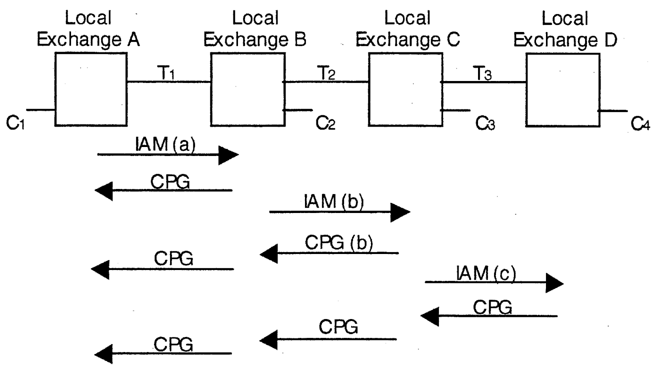


Figure 11.6-1. User-to-user signaling.



Contents of IAM			
Called Party Number (Par.6)	N2	N3	N4
Original Called Party Number (Par.20)	PNI	N2	N2
Redirecting Number (Par.21)	PNI	N2	N3
Redirection Information (Par.22)	PNI		
Number of Redirections		1	2

PNI: Parameter not included

Figure 11.6-2. Call forwarding.

the forwarding. If the calling customer is an ISDN user, the originating exchange sends this information to the customer in a Q.931 message.

11.6.6 Congestion at Exchange

The procedure is the same as in TUP (Section 9.5.1). When an exchange A is overloaded, it informs the concerned exchanges (exchanges with ISUP trunk groups to A) by including a Par.3 (automatic congestion level) in the REL (Release) messages sent for these trunks. Par.3 can indicate a moderate or a severe congestion.

On receipt of a congestion indication, the concerned exchanges temporarily reduce, or suspend, the seizures of trunks to exchange A.

11.6.7 Circuit Supervision

ISUP includes procedures and messages that request a circuit (trunk) to be blocked, unblocked, or reset, and messages that acknowledge those requests. They correspond to the TUP messages discussed in Section 9.5.2. The message acronyms and type codes are listed in Table 11.6-1. No parameters are included in these messages.

11.6.8 Circuit Group Supervision

The messages and procedures for blocking, unblocking, and resetting groups of trunks correspond to those in TUP (Section 9.5.4). The names, acronyms, and type codes of the messages are listed in Table 11.6-2. The message parameters are outlined below.

Range and Status. This is a mandatory parameter in all circuit group supervision messages. It consists of two subfields (Fig. 11.6-3). The *range* subfield contains a number R that indicates the number of circuits in the affected group. The affected circuits are specified by R (range) and CIC (circuit identification code). If CIC = c and R = r, the potentially affected circuits are those whose CICs are in the range from c through c + r.

In the ITU-T and U.S. versions of ISUP, the maximum values of R are 31 and 23, respectively. In this way, all trunks in an E1 (European) or T1 (North American)

TABLE 11.6-1 Circuit Supervision Messages

Acronym	Name	Message Type Code
BLA	Blocking Acknowledgment	0001 0101
BLO	Blocking	0001 0011
CCR	Continuity Check Request	0001 0001
RSC	Reset Circuit	0001 0010
UBA	Unblocking Acknowledgment	0001 0110
UBL	Unblocking	0001 0100

TABLE 11.6-2 Circuit Group Supervision Messages

Acronym	Name	Message Type Code
CGB	Circuit Group Blocking	0001 1000
CGBA	Circuit Croup Blocking Acknowledgment	0001 1010
CGU	Circuit Group Unblocking	0001 1001
CGUA	Circuit Group Unblocking Acknowledgment	0001 1011
GRS	Circuit Group Reset	0001 0111
GRA	Circuit Group Reset Acknowledgment	0010 1001

Source: Rec. Q.763. Courtesy of ITU-T.

first-order digital multiplex system (Section 1.5.2) can be blocked or unblocked simultaneously.

The *status sub-field* has variable length and has a status bit for the individual circuits. Bits 1, 2, ..., 8 in octet 2 represent circuits with CIC = c, c + 1, ..., c + 7; bit 1 in octet 3 represents the circuit with CIC = c + 8, and so on.

In the Circuit Group Blocking and Unblocking messages (CGB and CGU), a status bit with value "1" indicates that the corresponding trunk should be, respectively, blocked or unblocked. In Circuit Group Blocking and Unblocking Acknowledgment messages (CGBA and CGUA), a status bit with value "1" indicates that the requested blocking or unblocking of the corresponding trunk is now in effect.

A Circuit Group Reset Request (GRS) message includes only the range octet of the range and status parameter. When exchange B receives a GRS message with CIC = c and R = r, it releases all calls on circuits with CIC values c through c + r.

A Circuit Group Reset Acknowledgment (GRA) message includes a complete range and status parameter; the bits in the status field represent individual trunks in the group specified by the values of CIC and R. Exchange B sets the status bits of a trunk to "1" if it is currently blocking the trunk at its end. This informs exchange A of which trunks cannot be seized for new calls.

Circuit Group Supervision Message Type. This is a parameter that appears in the messages CGB, CGU, CGBA, and CGUA (Circuit Group Blocking, Unblocking, and their acknowledgments) only. The parameter has a one-octet value field, in which bits 2 and 1 indicate the reason for the blocking or unblocking.

Octets↓	8	7	6	5	4	3	2	1	←Bits
1	Range								
2	Status								
..									
n									

Figure 11.6-3. Range and status parameter.

Bits 2,1 = 0,0 Blocking for maintenance reasons. No trunk should be seized for new calls, but existing calls are allowed to continue.

Bits 2,1 = 0.1 Blocking because of a hardware failure. No new calls are allowed, and existing calls have to be released immediately.

The "hardware failure" blocking and reset requests for a circuit group usually interrupt a number of calls. Therefore, as a safety measure against an undetected error in a CGB or GRS message, an exchange repeats a sent CGB or GRS message within 5 s. The other exchange only carries out the blocking or reset procedure if it has received two identical messages within 5 s.

11.6.9 Signaling Route Set Congested or Unavailable

These procedures are the same as in TUP signaling (Section 9.5.5).

Signaling Route Set Congested. When the ISUP at an exchange receives a MTP-status indication for a destination (Section 8.9.3), it temporarily reduces the number of seizures of outgoing trunks in the routes to that destination.

Signaling Route Set Unavailable. On receipt of a MTP-pause indication for a destination (Section 8.9.5), ISUP clears all ISUP trunks to that destination and suspends seizing those trunks for new calls until it receives a MTP-resume indication.

11.6.10 Application Transport Mechanism (APM)

The application transport mechanism [7] is an ISUP procedure that can be used by two peer applications, residing in separate exchanges, to send application-specific data sequences to each other. The signaling relationship between applications established by the APM procedure is similar to what is provided by TCAP (Chapter 16). The data is encapsulated in the application transport parameter (Par.2—ATP) and transported transparently in the appropriate ISUP message. The APM procedure allows for segmentation and reassembly of data sequences that do not fit into a single message. The maximum length allowed for a data sequence is 10 segments and 2048 octets. It should be pointed out that the APM acronym is used to denote both the "mechanism" (procedure) and one of the message types that may be used to invoke it. The procedure may be invoked with any of the messages that can carry the ATP parameter (Table 11.2-2): it may use an established connection (e.g., by inserting the ATP parameter in the IAM message or by sending an APM message at any time during the call), or the application needing the procedure can start the connection.

Each message involved in the APM procedure can carry more than one instance of the ATP parameter (Par.2), each related to a different application, as indicated by the ACI field value.

11.6.11 Fallback

This feature allows a calling ISDN user to fall back to a simpler connection type when the preferred type cannot be handled.

Recommendation Q.764 [6] describes an example of a caller who prefers a 64-kb/s unrestricted connection but likes to fall back to a speech connection if the preferred connection cannot be made. In this case, the caller includes two bearer capability (BC) parameters in his ISDN SETUP message: BC to indicate the preferred connection type, and BC′ to indicates the fallback connection type. The originating exchange maps the parameters BC, BC′ into the requested TMR and USI and the fallback TMR′ and USI′, respectively. If all goes well, the connection is set up based on TMR. If an exchange along the path determines that it cannot extend the setup with this connection type, it deletes TMR and USI from its IAM and continues the setup based on TMR′. In this case, the caller must be warned. This is accomplished by including the parameter TMU, which indicates the actual connection type, in a backward message like ACM (Address Complete) or ANM (Answer).

Another example of a fallback scenario is when the called user has indicated in a Call Proceeding (CALPRC), Progress (PROG), or Alerting (ALERT) message that the call should fall back to speech. The destination exchange then sets up the call and alerts the calling user by including the parameter TMU = speech in its ACM (Address Complete message). The originating exchange then informs the calling user in its Alerting message.

There are several other fallback causes. For instance, the destination exchange may determine that the called party is not an ISDN user, or an intermediate exchange may know that the next exchange in the connection does not support fallback.

11.6.12 Multirate Connections

Multirate connections allow data transfer between ISDN users at rates that are multiples of 64 kb/s. Connections of 128, 384, 1536, and 1920 kb/s (2, 6, 24, and 30 time slots) were made available first, and N $\times$ 64-kb/s rates followed later.

In ISUP messages the parameter circuit identification code (CIC) identifies the channel (time-slot) number in 64-kb/s connections and the lowest numbered channel in multirate connections (Fig. 11.2-2). In multirate connections the number of 64-kb/s channels is indicated by Par.25 (TMR).

To extend a multirate call, the incoming exchange needs to know all channels in the connection. There are two methods to achieve that:

1. The outgoing exchange assigns consecutively numbered channels.
2. The outgoing exchange assigns channels at random.

With the first method, for example, channels 5, 6, 7, 8, 9, 10 may be assigned for a 384-kb/s connection. In this case the incoming exchange can identify all channels from the values CIC and TMR.

The second method increases the traffic capacity of the facility. As an example, consider an arriving connection with six channels. Assume that there are no six consecutive idle channels in the outgoing link, but that the total number of idle channels is at least six. Under consecutive assignment the setup fails, but under random assignment it is successful.

Random assignment is indicated by the outgoing exchange with the inclusion of a circuit assignment map (Par.10) in the IAM message.

11.7 SIGNALING PROCEDURES FOR FAILED SETUPS

The signaling procedures for interexchange calls that arrive on ISUP trunks at an exchange where a setup failure occurs depend on whether the calling party is an analog subscriber or an ISDN user, and on whether the connection that has been setup so far has ISUP signaling all the way. This section explores a number of setup failures.

11.7.1 Calls Originated by ISDN Users, ISUP Signaling All the Way

We first consider the case that the setup of a call originated by a user at exchange P has reached the destination exchange R, and that one of the failures described in Section 10.6.1 occurs. The called party can be a subscriber or a user. From Par.14 (forward call indicators) of the IAM, exchange R knows that the interexchange signaling is ISUP all the way, and that the calling party is an ISDN user.

In most failures, terminating exchange R does not have to provide in-band information to the calling TE. It simply initiates the release of the connection (Fig. 11.7-1) and sends a REL message that includes Par.9 (cause). The coding of

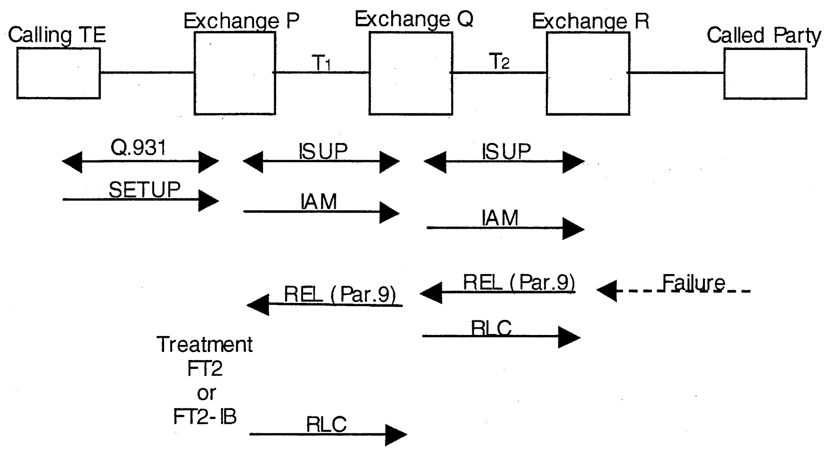


Figure 11.7-1. Failure procedure when exchange R does not provide in-band information.

this parameter is the same as that of Q.931 information element IE.6 (Section 10.3.4). Originating exchange P bases its treatment of the calling TE on the value of Par.9, which can indicate "user busy," "alerting, no answer," and so on.

On 64-kb/s calls, exchange P does not provide in-band information and the calling TE receives failure treatment FT2 (Section 10.6.1). On speech and 3.1-kHz audio calls, the treatment may require that exchange P provide in-band information (busy- or reorder-tone) to the calling TE (failure treatment FT2-IB).

On some failure of speech and 3.1-kHz audio calls, terminating exchange R provides ringing-tone, or an announcement, to the calling TE. Figure 11.7-2 shows the procedure for the case that the exchange has determined that the called number is not assigned. The exchange connects trunk T_2 to an announcement source and sends an ACM message that includes a Par.9 with location and cause values "public network serving the remote user" and "unassigned number." The message also includes a Par.18 (optional backward call indicators) with value "in-band tone or announcement now available."

Originating exchanges P then gives treatment FT2-IB to the calling TE and, when the TE responds with a DISC message, starts the release of the network connection.

11.7.2 Calls Originated by Subscribers, ISUP Signaling All the Way

If the calling party is a subscriber, the procedures at the terminating exchange R are, with minor exceptions, as described above for speech or 3.1-kHz audio calls. On calls where the calling subscriber should hear busy-tone or reorder-tone, exchange

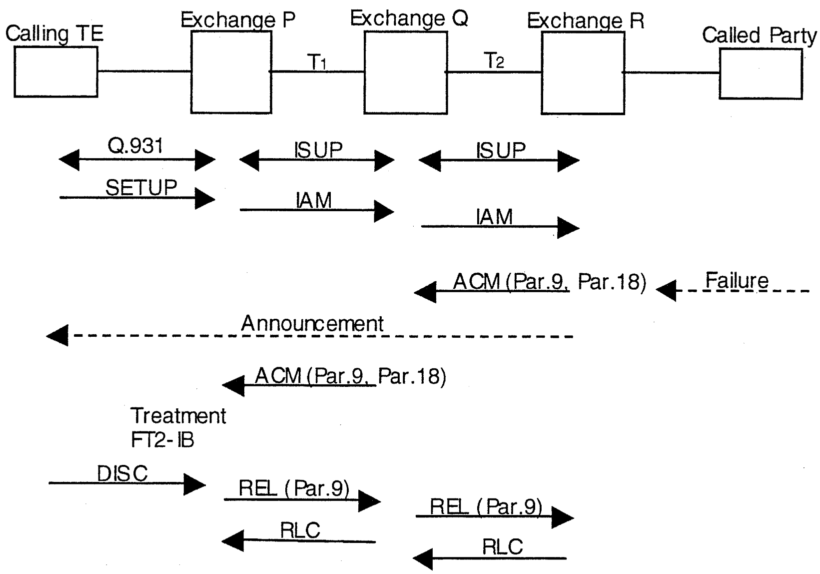


Figure 11.7-2. Unassigned number received by exchange R, on a speech or 3.1 kHz audio call.

R initiates the release of the connection. Exchange P provides the tone according to the information in Par.9 (cause) of the received REL message, releases trunk T_1 (as in Fig. 11.7-1), and awaits the disconnect signal from the calling subscriber. Ringing-tone or an announcement may be provided to the subscriber by terminating exchange R (as in Fig. 11.7-2). When exchange P receives the subscriber's disconnect signal, it initiates the release of the connection.

11.7.3 Failure at Intermediate Exchange, ISUP Signaling All the Way

Suppose that the setup of a call has arrived at intermediate exchange Q of Fig. 11.7-1, and that the setup fails at that exchange. Trunk T_1 has ISUP signaling and Par.14 (forward call indicators) in the IAM received by exchange Q thus indicates "ISUP signaling all the way."

On 64-kb/s calls, Q does not provide in-band information. It releases the connection, sending a REL message that includes Par.9 (cause indicators). Originating exchange P then sends a DISC message to the calling TE (treatment FT2), which includes an IE.6 (cause), in which the cause and general location information is copied from Par.9.

On speech and 3.1-kHz audio calls, exchange Q provides in-band information. It connects the incoming trunk to the appropriate in-band source and returns an ACM message that includes Par.9 and Par.18 (optional backward call indicators) indicating "in-band information now available."

Typical cause values and locations for failures that occur at intermediate exchanges are (Par.9):

Cause Value	
2	No route to specified transit network
3	No route to destination
34	No outgoing circuit (trunk) or channel available

Location	
2	Public network of local user
3	Transit network
4	Public network of remote user

11.7.4 Failures on Connections with Signaling Interworking

Finally, consider failures occurring at terminating exchange R in the connection of Fig. 11.4-2(*b*). From Par.14 (forward call indicators) in the IAM, exchange R knows that signaling interworking has occurred, and that the call must therefore be a speech or 3.1-kHz audio call. Exchange R then connects trunk T_2 to an announcement or tone source. When exchange P receives a DISC message or a disconnect signal from the calling ISDN user or subscriber, it initiates the release of the network connection.

11.8 ISUP SIGNALING IN THE INTERNATIONAL NETWORK

11.8.1 Introduction

In the mid-1980s, representatives from AT&T, British Telecom International, and KDD (Japan) started to define a version of ISUP for use in the international network [8]. International ISUP has been installed in international switching centers (ISCs) of many countries and is used for call control on a large number of international trunk groups.

International ISUP, as documented in Rec. Q.767 [9], is a subset of ITU-T Recommendations Q.762–Q.764. Address signaling is *en-bloc* only (the IAMs always include the complete called party number). End-to-end signaling and call forwarding are not available.

11.8.2 Service Supported by International ISUP

In addition to bearer services and the transfer of parameters for TE selection at the called ISDN user, international ISUP supports the following supplementary services: "calling line identification presentation/restriction," "connected line identification presentation/restriction," "user-to-user signaling," and "closed user group service." This support is limited to the transfer of parameters associated with these services across the international network.

11.8.3 Messages Not Provided

Of the messages listed in Sections 11.2.3, 11.6.7, and 11.6.8, international ISUP does not provide:

- Information (INF)
- Information Request (INR)
- Pass-Along (PAM)

11.8.4 Parameters Not Provided

Of the parameters described in Sections 11.2.4 and 11.6.8, international ISUP does not provide:

- Call reference (Par.5)
- Information indicators (Par.15)
- Information request indicators (Par.16)

11.8.5 Forward and Backward Call Indicators

Since international ISUP signaling does not support end-to-end signaling, bits H, G of Par.4 (backward call indicators) and bits C, B of Par.14 (forward call indicators) are always set to zero.

11.9 ISUP SIGNALING IN THE UNITED STATES

This section outlines ISUP signaling in U.S. networks. The U.S. version of ISUP has been standardized by the American National Standards Institute (ANSI) and Bellcore (now Telcordia) [10–13].

11.9.1 Services

The services supported by the U.S. version of ISUP are largely similar to those of ITU-T ISUP. They include call forwarding and all services that require the transfer of the calling party number to the terminating exchange.

Not included are user-to-user signaling for ISDN users and closed user group (CUG) service. However, U.S. ISUP supports CUG on international transit calls. For example, if an international switching center (ISC) on the West Coast receives an IAM for a trunk from Japan (Fig. 11.9-1), the call destination is in the United Kingdom, and the IAM includes a Par.11 and Par.19 (closed user group interlock code and optional forward call indicators), these parameters are included in the IAMs for the trunks in the United States and the IAM for the outgoing international trunk.

Some services and procedures provided in the United States, but not in other countries, are outlined next.

11.9.2 Business Group Services

In the United States a group of lines, or a private branch exchange (PBX), can be part of a *business group* (BG) [10]. Business groups can be established when requested

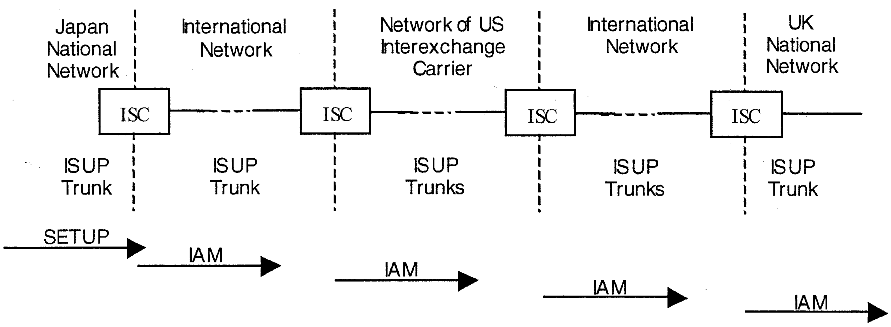


Figure 11.9.1. Call from Japan to the UK.

Octets↓	8	7	6	5	4	3	2	1	←Bits
1	Spare	Atten-dant Status	BGID	LP	Party selector				
2	MSB								
3	Business Group identifier								
4								LSB	
5	MSB								
	Sub-group identifier								
6								LSB	
7	Line privileges								

MSB: Most significant bit LSB: Lease significant bit

Figure 11.9.2. Business group parameter. (From ANSI T1.113-1992. Reproduced with permission of the Alliance for Telecommunications Industry Solutions, Inc.)

by a business customer, a government agency, and so on. Business groups have some similarities with closed user groups (CUGs). At the request of individual business customers, the network can restrict calls to/from the lines of their BGs. A local exchange stores data about each attached BG line.

On calls originated by a BG line, those data are also needed at the terminating exchange. On these calls, the originating exchange therefore includes a "business group" parameter in its IAM (Fig. 11.9-2).

The *attendant status* bit indicates whether the line is an attendant line (1) or not (0). The BGID bit is set to 0 for multilocation BGs, and to 1 if the BG is part of a private (corporate) network. The *party selector* indicates whether the parameter applies to the called, calling, original called, or redirected number in the IAM. The *BG identifier* is a binary number that represents the BG. A business customer can also allocate subgroup identifiers, for example, to individual locations of the BG.

The *LP bit* indicates whether the privileges of the line are network-defined (0), or defined by the BG customer (1). For network-defined privileges, bits 8 through 5 and 4 through 1 specify restrictions on originating and terminating privileges, respectively:

0000	Unrestricted
0001	Semirestricted
0010	Fully restricted, intraexchange
0100	Privileges denied

11.9.3 Transit Network Selection

Long distance calls in the United States are handled by an interexchange carrier (IXC) selected by the calling party (Section 3.7.1). If the caller's local exchange does not have a direct trunk group to an exchange in the IXC network, the IAM

8	7	6	5	4	3	2	1
Spare	Type of network identification			Network identification plan			
Digit 1				Digit 2			
Circuit code				Digit 3			

8	7	6	5	4	3	2	1
Spare	Type of network identification			Network identification plan			
Digit 2				Digit 1			
Digit 4				Digit 3			
Circuit code				Reserved			

Figure 11.9.3. Transit network selection parameter. (From ANSI T1.113-1992. Reproduced with permission of the Alliance for Telecommunications Industry Solutions, Inc.)

sent by the originating exchange includes a transit network selection [10] parameter that identifies the IXC. The contents of the parameter are shown in Fig. 11.9-3 [10].

The *type of network identification* field is set to 010. That means that the identification is in accordance with a national network standard (as opposed to an ITU-T standard). IXC networks have codes consisting of three or four digits. This is indicated by the *network identification plan* field:

0001 Three-digit identification code
0010 Four-digit identification code

The *circuit code* field shows whether the call is international and whether operator assistance has been requested:

0000 National call
0001 International call, no operator requested
0010 International call, operator requested

The parameter is removed from IAM by the exchange that seizes an outgoing trunk to an exchange in the network of the selected IXC.

11.9.4 Other Differences

In conclusion, we point out some messages and procedures that are particular to United States ISUP:

- All address signaling is *en-bloc*; subsequent address messages are not used.
- In IAMs of ITU-T ISUP, Par.27 (user service information) is optional (included only on calls originated by ISDN users). It holds the bearer capability and is

transferred transparently by the network. Mandatory Par.25 (transmission medium requirement) is used by the network for the selection of outgoing trunks of the appropriate type (e.g., digital trunks for 64-kb/s data calls). In the United States, Par.25 is not used. Instead, Par.27 is mandatory and in calls originated by subscribers it is set to "3.1-kb/s audio." Par.27 is examined by the exchanges along the connection to determine the type of outgoing trunk and is mapped into IE.1 (bearer capability) if the called party is an ISDN user.

- U.S. long-distance calls involve the local networks of the calling and called parties and an IXC network. During the setup of a call, the exchange in the originating local network that has seized an outgoing trunk to an exchange in the IXC network sends a backward Exit message [10]. The message, which is repeated all the way to the originating exchange, confirms that the setup in the originating network is complete. Likewise, when an IXC exchange has seized an outgoing trunk to an exchange in the terminating local network, it sends an Exit message, which is repeated to all IXC exchanges along the connection.
- In addition to the Par.9 (cause indicators) values standardized by ITU-T (Section 11.2.4), U.S. ISUP includes a number of values that have been specified by ANSI [10]. Some examples of cause values are:

001 0111	Unallocated destination number
001 1000	Unknown business group
011 1110	Call blocked by business group restrictions

11.10 ACRONYMS

ACI	Application context identifier
ACM	Address Complete message (ISUP)
ALERT	Alerting message (Q.931)
ANM	Answer message (ISUP)
ANSI	American National Standards Institute
APM	Application Transport message, application transport mechanism
ATP	Application transport parameter
BC	Bearer capability parameter (Q.931)
Bellcore	Bell Communications Research
BG	Business group
CALPRC	Call Proceeding message (Q.931)
CCIS	Common-Channel Interoffice Signaling
CGB	Circuit Group Blocking message
CGBA	Circuit Group Blocking Acknowledgment message
CGU	Circuit Group Unblocking message
CGUA	Circuit Group Unblocking Acknowledgment message
CIC	Circuit identification code

CID	Call identity
CLASS	Custom local area subscriber services
CON	Connect message (ISUP)
CONACK	Connect Acknowledgment message (Q.931)
CONN	Connect message (Q.931)
COT	Continuity message (ISUP)
CPG	Call Progress message (ISUP)
CUG	Closed user group
DISC	Disconnect message (Q.931)
DPC	Destination point code
DSL	Digital subscriber line
FDM	Frequency-division multiplexing
FOT	Forward Transfer message (ISUP)
GRA	Circuit Group Reset Acknowledgment message
GRS	Circuit Group Reset Request message
IAM	Initial Address message (ISUP)
IXC	Interexchange carrier
IE	Information element (in Q.931 messages)
INF	Information message (ISUP)
INR	Information Request message (ISUP)
ISC	International switching center
ISDN	Integrated Services Digital Network
ISUP	ISDN User Part
IXC	Interexchange carrier
LATA	Local access and transport area
MF	Multifrequency signaling, mandatory fixed parameter
MTP	Message Transfer Part (SS7)
MV	Mandatory variable-length parameter
NPIBS	Network-provided in-band information service
OP	Optional parameter
OPC	Originating point code
PAM	Pass-Along message (ISUP)
Par.	Parameter (in ISUP messages)
PC	Point code
PCI	Protocol control indicator
PCM	Pulse code modulation
RCI	Release call indicator
REL	Release message (ISUP)
RES	Resume message (ISUP)
RL	Routing label
RLC	Release Complete message (ISUP)
RLCOM	Release Complete message (Q.931)
RLSE	Release message (Q.931)
SAM	Subsequent Address message

SCCP	Signaling Connection Control Part
SETACK	Setup Acknowledgment message (Q.931)
SETUP	Setup message (Q.931)
SI	Service indicator, sequence indicator
SIO	Service indicator octet
SLR	Segmentation local reference
SLS	Signaling link selector
SNI	Send notification indicator
SS7	Signaling System No. 7
SSF	Subservice field
SSN	Subsystem number
SUS	Suspend message (ISUP)
TCAP	Transaction Capabilities Application Part
TE	Terminal equipment (ISDN user)
TMR	Transmission medium requirements
TMU	Transmission medium used
TUP	Telephone User Part
UM	User message
USI	User service information
UUI	User-to-user information

11.11 REFERENCES

1. A. R. Modaressi and R. A. Skoog, Signaling System No. 7: A Tutorial, *IEEE Commun. Mag.*, **28**, No. 7, July 1990.
2. R. Manterfield, *Common Channel Signalling*, Chapter 6, Peter Peregrinus Ltd., London, 1991.
3. *Signalling System No. 7—ISDN User Part Functional Description*, Rec. Q.761, ITU-T, Geneva, Dec. 1999.
4. *Signalling System No. 7—ISDN User Part General Functions of Messages and Signals*, Rec. Q.762, ITU-T, Geneva, Dec. 1999.
5. *Signalling System No. 7—ISDN User Part Formats and Codes*, Rec. Q.763, ITU-T, Geneva, Dec. 1999.
6. *Signalling System No. 7—ISDN User Part Signalling Procedures*, Rec. Q.764, ITU-T, Geneva, Dec. 1999.
7. *Signalling System No. 7—Application Transport Mechanism*, Rec. Q.765, ITU-T, Geneva, June 2000.
8. J. J. Lawser et al., Common Channel Signalling for International Services, *ISS Conf. Record*, Stockholm, 1990.
9. *Application of the ISDN User Part of CCITT Signalling System No. 7 for International ISDN Interconnections*, Rec. Q.767, ITU-T, Geneva, Feb. 1991.

10. *American National Standard for Telecommunications—Integrated Services Digital Network (ISDN) User Part*, ANSI T1.113-1992, American National Standards Institute, New York.
11. *ISDN Access Call Control Switching and Signaling Requirements*, TR-TSY-000268, Issue 3, Bellcore, Morristown, NJ, 1989.
12. *Switching System Generic Requirements for Call Control Using the Integrated Services Digital Network User Part (ISDNUP)*, Tech. Ref. TR-NWT-000317, Issue 4, Bellcore, Morristown, NJ, 1993.
13. *Switching Systems Generic Requirements Supporting ISDN Access Using the ISDN User Part*, Tech. Ref. TR-NWT-000444, Issue 3, Bellcore, Morristown, NJ, 1993.

12

SIGNALING IN CELLULAR MOBILE TELECOMMUNICATIONS

Cellular mobile telecommunications is one of the most important telecommunication developments of the last decade. The technical concepts underlying this type of communications were developed by Bell Laboratories [1,2] and implemented in the *advanced mobile phone service* (AMPS) *system*.

Early mobile stations (MSs) were designed as car phones. They were too bulky to be carried around and had to be powered by the battery of the car. Today, there are compact lightweight MSs with internal rechargeable batteries. They can be carried by hand and are "personal" phones rather than car phones.

There are two groups of signaling procedures in cellular mobile telecommunications. This chapter describes the signaling between a MS and a cellular mobile network. The second group of signaling procedures involves various entities in a mobile network and is discussed in Chapter 19.

Sections 12.1–12.6 of this chapter cover signaling in the AMPS system and its successors in the United States. Sections 12.7–12.9 describe signaling in the *global system for mobile telecommunications* (GSM), which has emerged as the most important cellular system outside the United States. Systems based on CDMA technology are discussed in Chapter 13.

12.1 INTRODUCTION TO CELLULAR MOBILE NETWORKS

This section briefly describes some important aspects of cellular mobile networks [3–5].

Signaling in Telecommunication Networks, Second Edition, by John G. van Bosse and Fabrizio U. Devetak

12.1.1 Definitions

A *cellular mobile network* (CMN) provides communication services for mobile stations that are operating in its *service area*. A service area typically covers a metropolis and its surrounding suburbs, or a number of medium-sized cities. The size of a service area is typically in the range of 100–4000 square miles.

The CMN service area is divided into a number of *MSC areas*, each of which contains an exchange known as a *mobile switching center* (MSC)—see Fig. 12.1-1. A MSC provides service to all MSs in its area. Some rural CMNs consist of just one MSC area. A MSC area is divided into a number of *location areas*, and each of these areas is divided into a number of *cells*. The cells are approximately circular, with radii that range from about 2 to 15 miles. Each cell has a *base station* (BS)—also known as *land station* and *cell site*—which houses radiofrequency (RF) transmitters and receivers.

A MSC has trunk groups (TGs) to nearby exchanges in the public switched telecommunication network (PSTN) (also known as the "fixed" or "wireline" network) and a *base station trunk group* (BSTG) to each base station—Fig. 12.1-2(*a*). When a CMN has several MSCs, there are also trunk groups (MSCTGs) between these MSCs—Fig. 12.1-2(*b*).

A MS operating in a cell communicates on a RF channel with the BS of the cell. There are two channel types: voice channels and control channels.

Voice Channels. In a base station, each BS trunk is permanently wired to the transmitter and receiver of a RF voice channel. The combination of a BS trunk and its associated voice channel is the functional counterpart of a trunk in the PSTN. A BS trunk and voice channel is assigned to a mobile at the start of a call and released when the call ends. Figure 12.1-3 shows the connection for a call

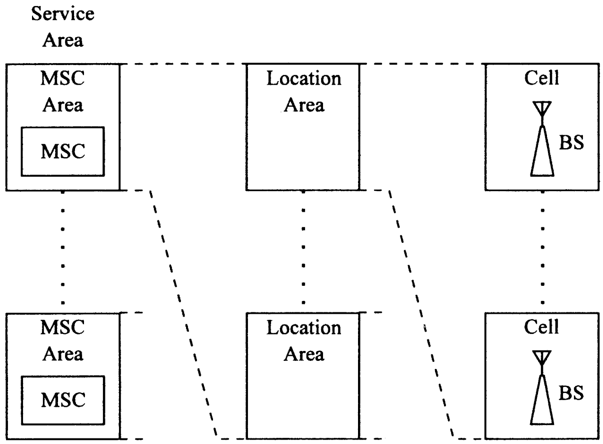


Figure 12.1-1. Cellular mobile network. (From EIA/TIA 553. Reproduced with permission of TIA.)

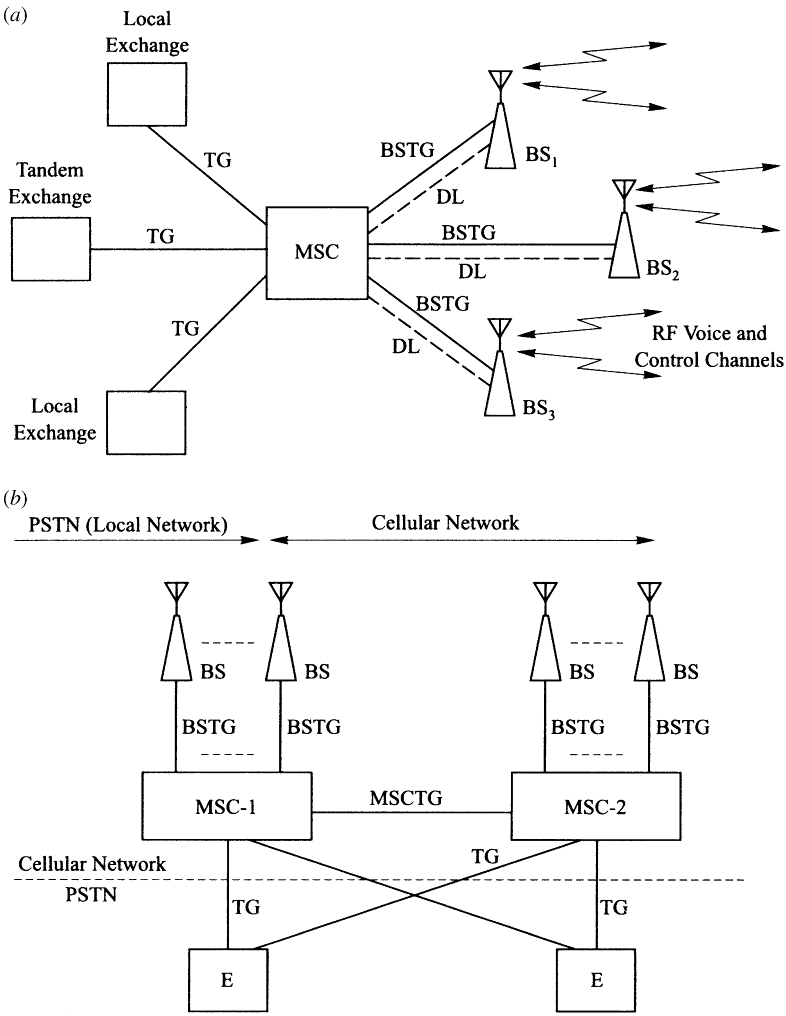


Figure 12.1-2. Mobile network equipment.

between a PSTN subscriber (S) and a mobile station (MS) operating in the cell of BS_1. The connection involves a path in the PSTN between S and exchange E, trunk T, a path—in the switchblock of MSC—between T and base station trunk (BST), trunk BST, and its associated RF voice channel (VC).

Control Channels. A base station has a pair (for reliability) of bidirectional data links (DLs) to its MSC (Fig. 12.1-2(*a*)). In the BS, the data links are connected to the RF equipment of a group of RF channels, known as *control channels*. These channels carry signaling messages to and from mobiles that are operating in the cell when

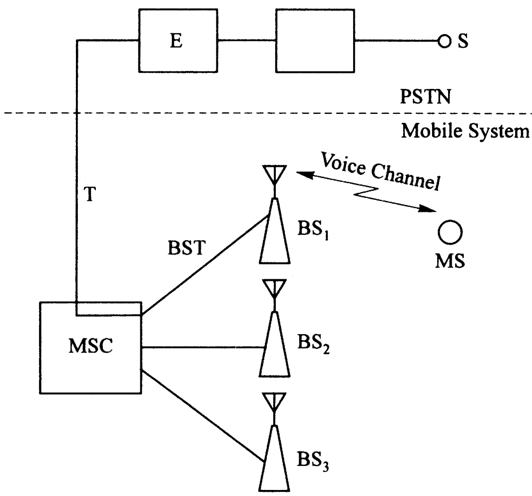


Figure 12.1-3. Connection between mobile (MS) and subscriber (S).

they are idle (turned "on" but not involved in a call). When a mobile is on a call, the voice channel that has been allocated to the call carries both speech and signaling.

We distinguish three types of control channels. *Paging* channels carry "paging" messages that are sent from the MSC to inform a mobile that it is being called. *Access* channels are used primarily by mobiles to originate a call and to respond to a received paging message. *Combined* control channels are used for paging and accessing. A MSC is equipped either with separate paging and access channels or with combined control channels.

12.1.2 AMPS Radiofrequency Channels

In mobile cellular literature, a channel is a bidirectional RF transmission facility, consisting of a *forward channel* for transmission from BS to MS and a *reverse channel* for transmission in the opposite direction.

The forward and reverse channels of AMPS are analog channels, spaced at 30-kHz intervals. On the voice channels, speech is transmitted by frequency modulation. The digital messages on the voice and control channels are transmitted by frequency-shift keying, with a signaling speed of 10 kb/s.

The Federal Communication Committee, which controls the use of the RF spectrum in the United States, originally allocated 40 MHz in the 850-MHz band to cellular mobile communications. This allowed 666 bidirectional channels in a service area. In 1987, the cellular spectrum was increased to 50 MHz, to accommodate 832 channels. Two competing cellular carriers—denoted as the "A" and the "B" carriers—are allowed to operate in a service area, using 416 channels each (395 speech channels and 21 control channels).

The 30-kHz channels are identified by *channel numbers*, which have been assigned as follows:

	"A" System	"B" System
Control channels	313–333	334–354
Voice channels	001–312 and 667–716 and 991–1023	355–666 and 717–799

The center frequency (f_c) of a forward channel can be determined from its channel number (N):

$$\text{For } N \text{ from 1 through 799:} \quad f_c = (0.03\,N + 870)\text{ MHz}$$

$$\text{For } N \text{ from 991 through 1023:} \quad f_c = (0.03\,N + 839.31)\text{ MHz}$$

The center frequencies of the reverse channels are 45 MHz below the center frequencies of their associated forward channels. When a MS is communicating on a particular channel, its receiver and transmitter are tuned to the center frequencies of the forward and reverse channel, respectively.

12.1.3 Frequency Reuse

A key concept in cellular mobile systems is *frequency reuse*: the same RF band (channel) is used in several cells of a cellular network. Without reuse, the maximum number of simultaneous calls in a CMN would be 395. Reuse greatly increases this number.

Frequency reuse is possible because, all other things being equal, the power of a received signal is roughly proportional to d^{-4}, where d is the distance between the transmitter and the receiver [3].

One can therefore allocate the same channels (frequencies) to cells that are sufficiently far apart from each other—say, five cell radii (R). In Fig. 12.1-4, mobile MS_1 is operating in cell 1 and communicating on channel N with BS_1. MS_2 in cell 2 is communicating with BS_2 on the same channel. MS_1 receives a desired signal from BS_1 and an interfering signal from BS_2. The minimum distance between MS_1 and BS_2 is $4R$, while the maximum distance between MS_1 and BS_1 is R. The transmit power at the base stations is equal. Therefore, the desired signal is approximately $4^4 = 256$ times stronger than the interfering signal. In frequency-modulated RF channels, this ratio of signal strengths is sufficient to suppress the effects of cochannel interference caused by BS_2. For the same reason, the cochannel interference caused by BS_1 does not affect MS_2. Mobiles MS_1 and MS_2 also transmit with equal power, and the effects of cochannel interference, from MS_1 to BS_2 and from MS_2 to BS_1, cause no problems at the base-station receivers.

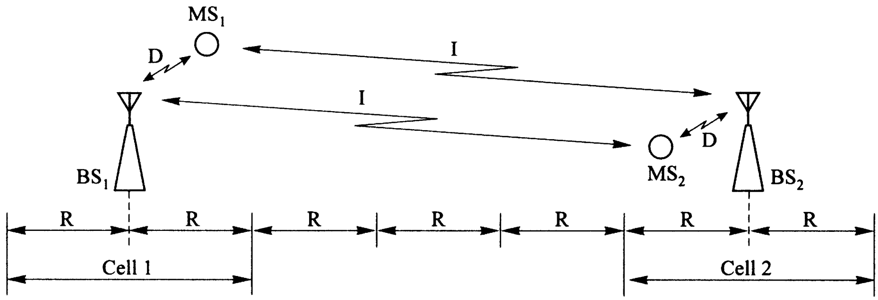


Figure 12.1-4. Cochannel interference D, Desired RF signals; I, interfering RF signals.

A frequently used cell configuration is shown in Fig. 12.1-5. The cells are shown as hexagons for convenience only, and actual coverage areas of adjacent cells overlap slightly. Seven adjacent cells (e.g., cells A_1 through A_7) form a cluster. The 395 voice channels and 21 control channels are divided into seven channel groups of about 56 voice channels and 3 control channels each. Cells A_1, $B_1, \ldots,$ G_1 use channel group 1; cells A_2, $B_2, \ldots,$ G_2 use channel group 2, and so on. In this way, there is no cochannel interference among cells of the same cluster.

Now consider cell A_1. The nearest cells that use the same channel group and could cause cochannel interference are B_1, $C_1, \ldots,$ G_1. When the cell radius equals R, the minimum distance between two potentially interfering base stations in this configuration is $4.6R$, which is sufficient in actual systems, except during severe fading of the desired signal.

12.1.4 AMPS Color Codes and Supervisory Audio Tones

RF signals are subjected to short fades (decreases in received signal power), for example, when the MS moves through the "RF shadow" of a building or because

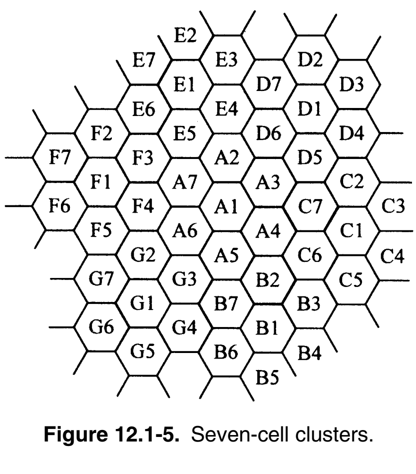


Figure 12.1-5. Seven-cell clusters.

of multiple reflections. Suppose that mobile MS is in cell A_1 of Fig. 12.1-5 and is using voice channel N of the base station of that cell. During a fade, the strength of channel N received by MS from cell A_1 may fall below the strength of channel N received from cell B_1. The user of the mobile would then hear a short burst of speech from the call in cell B_1.

AMPS uses *color codes* and *supervisory audio tones* (SATs) to provide protection against problems of this nature. A color code is assigned to each cell. The codes can have three values—00, 01, or 10—and each value is associated with a particular SAT frequency. Voice channels carry both speech and a SAT frequency.

Now consider the cells $A_1, B_1, \ldots, G_1$ of Fig. 12.1-5, all of which use the same channel group. The color codes and SAT frequencies could then be assigned as follows:

Color Code	SAT Frequency	Cells
00	5970 Hz	A_1
01	6000 Hz	B_1, D_1, F_1
10	6030 Hz	C_1, E_1, G_1

In this way, the color code of cell A_1 is different from the color codes of the cells in the adjacent clusters that use the same channel group and could cause cochannel interference. Also, cells B_1 and D_1, who are nearby cochannel interferers for cell C_1, have color codes that are different from cell C_1, and so on.

All base stations transmit the SAT frequency of their respective cells on their forward voice channels. A mobile operating in A_1 expects to receive—and usually receives—the 5970-Hz SAT frequency of cell A_1.

However, during a fade it may receive a forward channel from an interfering BS in cell $B_1, \ldots,$ or G_1. When receiving an unexpected SAT frequency, the mobile mutes the received speech. The mobile user then experiences a silent interval, instead of a more disturbing burst of extraneous speech.

All mobiles transmit the received SAT frequency on their reverse voice channels. A base station therefore expects to receive the SAT frequency of its cell on its reverse voice channels. When a BS receives a different SAT frequency, it mutes the received speech, and the subscriber connected to a mobile is therefore also protected against extraneous speech.

The SAT frequencies are well above the highest transmitted speech frequency. The MS and BS receivers separate the speech and the SAT frequency with low-pass and high-pass filters, and no SAT is heard by the listener at the MS or by the party at the other end of the connection.

12.1.5 Cell Size

With seven-cell clusters, each cell can have a group of up to 56 voice channels. The load (the average number of simultaneous calls) on a group of 56 channels should not exceed 42 erlangs. Otherwise, the probability that all channels are busy when a new call has to be set up becomes unacceptably high. When the traffic density in a mobile system is T erlangs per square mile, and A is the coverage area of the cell (square miles), we have $A < 42/T$. Assuming that the cell is approximately circular, its maximum radius R (miles) is

$$R^2 < \frac{42}{(3.14)T}$$

For example, with $T = 2$, R should be <2.6 miles, and therefore $A < 21$ miles^2. Call density T is usually high in metropolitan areas and low in rural areas. Therefore, metropolitan systems usually have a large number (say, up to some 100) of small cells, and rural systems have a small number of large cells, which may be equipped with less than 56 voice channels.

12.1.6 AMPS Transmitter Power Levels

Consider a cell with a radius R, with a BS located at its center. The maximum distance between the BS and a mobile in the cell then equals R. In order to have a prescribed minimum signal power level at the receivers in the mobile and base stations when the distance between the BS and MS is at its maximum (R), the transmit power for a channel has to be approximately proportional to R^4. The power level of a BS transmitter can be set when it is installed, because cell radius R does not change. However, mobiles can operate in cells with different radii R, and their transmit power has to be adjusted accordingly.

This is done on command from the MSC that is serving the mobile. Each BS periodically measures the signal strength received on its reverse voice channels and reports the results to its MSC. When the MSC decides that a MS needs to change its transmit power, it sends a "change power" message that includes a *mobile attenuation code* (MAC) whose value ranges from 0 through 7 and represents the requested power level: MAC $= 0$ requests $+6$ dBW, MAC $= 1$ requests $+2$ dBW, and each next higher value reduces the requested power by another 4 dB.

Most of the power in a mobile station is consumed by its transmitter. The transmitters of mobile stations fall in power class I, II, or III, which have maximum power levels of $+6$, $+2$, and -2 dBW, respectively. Car-mounted mobile stations have class I transmitters. The power in hand-held stations has to be used sparingly to avoid frequent battery recharging, and these stations usually have class III transmitters. When a mobile receives a command for a power level that exceeds the capability of its transmitter, it simply transmits at its maximum power. Class III transmitters sometimes cause problems when the mobile operates near the edge of a large cell.

Some hand-held mobiles have a power-saving feature known as *discontinuous transmission* (DTX). When a mobile is in a call and its user is talking, the transmit power is as described above, but when the user is silent, the transmit power is reduced, usually by 8 dB.

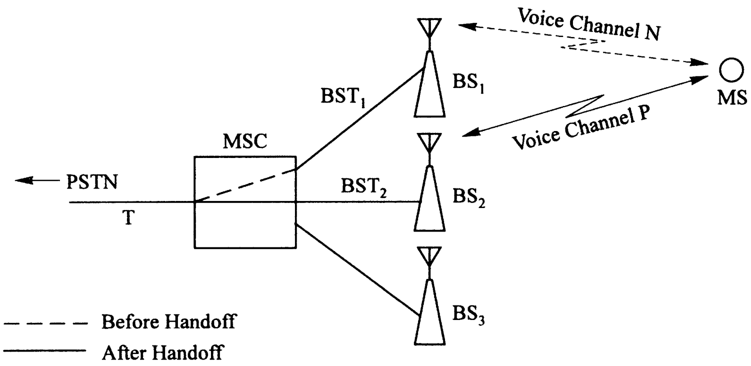


Figure 12.1-6. Handoff of a mobile.

12.1.7 Handoff

During a call, a mobile can move from one cell to an adjacent cell [3,6,7]. Suppose that in Fig. 12.1-6, mobile MS has started a call in the cell of BS_1. The connection involves trunk T to the PSTN, base-station trunk BST_1, and the associated voice channel N. When MS has moved to the cell of BS_2, it becomes necessary to hand it off to that base station.

The handoff decision is made by the MSC, which periodically requests BS_1 to measure the strength of the signal received from the mobile (on reverse channel N). When the strength falls below a predetermined level, the MSC requests the base stations in adjacent cells to measure the signal strength received on channel N. In this example, BS_2 reports the strongest signal and is, therefore, in the best position to serve MS. The MSC then seizes an idle voice channel P of BS_2, and the associated trunk BST_2. It then releases channel N and trunk BST_1, connects T with BST_2, and sends a "handoff" message that identifies the new channel P. The MS then tunes its transmitter and receiver to that channel.

12.1.8 Relationships of Mobile Station and Network

In the United States, each cellular mobile service area is covered by two competing CMNs, known as the "A" and the "B" network. The owner of a MS selects one of these as his service provider. The selected network is known as the "home" CMN of the mobile.

When the MS operates in a location outside the service area of its "home" CMN, it can receive service from the A or B network that covers its current location. Also, when the mobile is in the service area of its "home" CMN and is unable to access the "home" CMN, it can access the competing CMN.

When a mobile is receiving service from a CMN other than its "home" CMN, we say that the mobile is "roaming" and is served by a "visited" CMN.

12.2 AMPS TONE SIGNALS AND MESSAGE WORDS

This section begins the description of the AMPS signaling protocol. The protocol is based on the original development by Bell Laboratories [1,2] but includes a number of additions and modifications specified by the EIA/TIA (Electronics Industries Alliance/Telecommunications Industry Association)—see [6].

The signaling between mobile and base station is a combination of common-channel and channel-associated digital signaling messages, and a single-frequency signal.

The forward and reverse control channels (FOCC and RECC) carry signaling messages only. They are "common" channels, used for signaling between a base station and all mobiles in the cell that are active, but not involved in a call. The messages on these channels include a *mobile identification number* (MIN) that identifies a particular mobile.

A voice channel is allocated to a MS at the start of a call. Its forward and reverse voice channels (FVC and RVC) carry speech and channel-associated signaling messages (which do not include MIN).

12.2.1 Supervision Tone

This 10-kHz tone is sent on RVC only and represents the on-hook (tone on) and off-hook (tone off) states of the mobile. When the mobile is in a conversation, on-hooks of 400 ms and 1.8 s indicate "flashes" and "disconnects," respectively.

12.2.2 Transmission of Messages

Messages consist of one or more words, transmitted at 10 kb/s. On forward channels (FOCC, FVC), the word length is 40 bits (28 information bits, followed by 12 parity bits for error checking). Words on reverse channels (RECC, RVC) have 48 bits (36 information bits, and 12 parity bits). Figure 12.2-1(*a*), (*b*), and (*c*) show the general structure of transmitted word blocks for messages with two words on, respectively, RVC, FVC, and RECC. In the figure, the fields in the message streams are denoted by acronyms, and the corresponding field lengths (number of bits) are shown below the acronyms.

Repeated Transmission. Each message word is repeated several times. This greatly reduces the probability that a word is completely missed by a receiver because of fading. Figure 12.2-1(*a*) shows that message words W1 and W2 on a RVC are repeated five times (W1-1, ..., W1-5; W2-1, ..., W2-5). Words on FOCC and RECC (Fig. 12.2-1(*c*), (*d*)) are also repeated five times, but words on FVC (Fig. 12.2-1(*b*)) are repeated 11 times, for reasons to be discussed later.

Error Checking. Error correction by retransmission (used on SS7 signaling links—see Sections 8.4 and 8.5) is not practical in mobile signaling. Errors in received messages are minimized by taking advantage of the repeat transmissions

DOT	WS	W1-1	DOT	WS	W1-2	...	DOT	WS	W1-5	DOT	WS	W2-1	...	DOT	WS	W2-5
101	11	48	37	11	48		37	11	48	37	11	48		37	11	48

(*a*) Reverse Voice Channel (2-word Message)

DOT	WS	W1-1	DOT	WS	W1-2	...	DOT	WS	W1-11	DOT	WS	W2-1	...	DOT	WS	W2-11
101	11	40	37	11	40		37	11	40	101	11	40		37	11	48

(*b*) Forward Voice Channel (2-word Message)

DOT	WS	CODED DCC	W1-1	W1-2	...	W1-5	W2-1	...	W2-5
30	11	7	48	48		48	48		48

(*c*) Reverse Control Channel (2-word Message)

DOT	WS	WE-1	WO-1	WE-2	WO-2	...	WE-5	WO-5	etc.
10	11	40	40	40	40		40	40	

BI	DOT	BI	WS	BI	WE-1	BI	WE-1	BI	WE-1	BI	WE-1	BI
1	10	1	11	1	10	1	10	1	10	1	10	1

(*d*) Forward Control Channel

Figure 12.2-1. Transmission of message words.

of individual words. A MS or BS receiver first takes a majority vote of the corresponding bits in the N appearances of a word. It then checks the resulting "best guess" word for errors, using the parity (P) bits.

The sender of a message expects an acknowledgment in the form of a signal or a confirmation message. A failure to receive an acknowledgment within a specific time interval indicates that a problem has occurred.

Synchronization. Each message starts with a 10101... *dotting sequence* (DOT) that is used by the receiver for bit synchronization. The length of DOT depends on the channel type. DOT is followed by an 11-bit *word synchronization pattern* (WS): 11100010010. In voice-channel messages, DOT-WS sequences also appear between the message words.

Blank-and-Burst. Message on voice channels are sent as short data bursts (less than 0.1 s). While sending a message, the transmitter and receiver blank out the speech and tone signals.

Data Streams on Forward Control Channels—Fig. 12.2-1(d). FOCCs continually transmit three interleaved data streams. The streams of message words WO and WE are read by mobiles with, respectively, odd and even MINs. Each word is again repeated five times (WO-1, ..., WO-5, ..., etc.). The third stream consists of busy idle (BI) bits that appear at the start of the DOT and WS sequences, and before the 1st, 11th, 21st, and 31st bit of each word. They indicate the status of the associated reverse control channel:

BI	RECC Status
0	Busy (receiving a message from a mobile)
1	Idle

A RECC is a common channel for messages sent by idle mobiles within a certain cell or group of cells. To avoid "collisions" (simultaneous messages from two or more mobiles on the channel), a mobile that needs to send a message on a RECC first examines the BI bits on the associated FOCC. When the channel is idle, the mobile seizes the associated RECC and starts to transmit its message. A mobile that finds the RECC busy (receiving a message from another MS) waits a random interval (0–100 ms) before trying again.

12.3 INTRODUCTION TO AMPS SIGNALING

This section describes a number of basic signaling procedures between the mobile and the cellular mobile network (CMN). The focus is on the messages in the radio-frequency channels between a mobile (MS) and a base station (BS). The division of functions between the mobile switching center (MSC) and its base stations is implementation-dependent. We assume here that all "logic" of the CMN resides in the MSC, and that the BS merely transmits messages as directed by the MSC and reports all received messages to the MSC. At this point, only a small number of message parameters are discussed.

Channel Number (CHAN). This identifies a voice channel.

Voice-Channel Mobile Attenuation Code (VMAC). This indicates the power level at which the mobile should transmit on the voice channel.

System Identification (SID). This identifies a particular mobile system. By convention, the "A" and "B" cellular systems have odd and even SIDs, respectively.

Mobile Identification Number (MIN). A 10-digit national number that identifies a mobile. In the United States, the numbering plan for mobile network is integrated into the PSTN numbering plan. A MIN consists of a three-digit area and exchange codes AC-EC, followed by a four-digit "line number" LN (Section 1.2.1). The AC-EC of a MIN identify the "home" MSC of the MS. Calls to a MS are routed by the PSTN to its home MSC.

Mobile Serial Number (MSN) or Electronic Serial Number (ESN). Uniquely identifies a mobile station.

A mobile station has variable, semipermanent, and nonalterable memory devices. The SID of the "home" cellular system and the MIN assigned to the mobile are entered into semipermanent memory by an agent of the "home" cellular system. The MSN of a mobile station is assigned by the manufacturer and is stored in nonalterable memory.

12.3.1 Initialization

When a mobile is turned on, it has to establish contact with a cellular network. A mobile whose home is an "A" system first tries to establish contact with the "A" system that serves the area where the MS is located. If this fails, it tries the "B" system. Mobiles with "B" home systems first try to contact the "B" system.

In the example of Fig. 12.3-1, we assume that the mobile has an "A" home system. It starts by scanning the 21 dedicated forward control channels of "A" systems and tunes to the strongest one (i.e., a control channel transmitted by the nearest BS).

All forward control channels of a CMN broadcast *overhead parameter* messages, at intervals of 0.8 s. These messages contain system-specific parameters. A mobile has to acquire this information before it can access (send messages to) a cellular system. Examples of system-specific information are the SID of the CMN, data about its access and paging channels, and indicators that represent the characteristics and capabilities of the system.

When the mobile receives its first overhead message (a), it stores the received SID, and information about paging channels, in its variable memory. By comparing

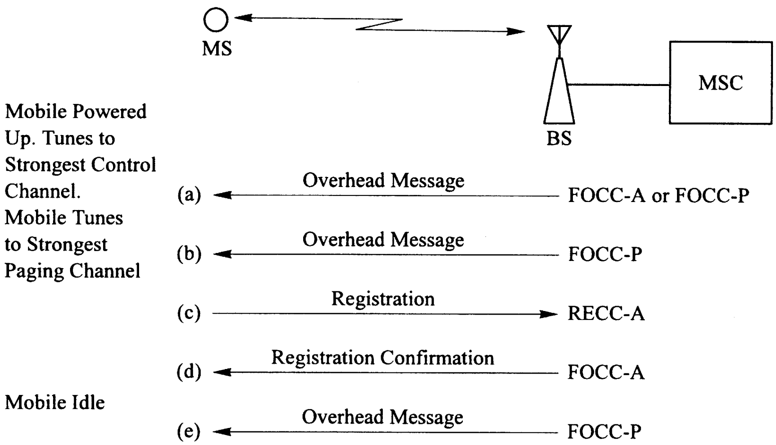


Figure 12.3-1. Initialization and registration. FOCC-A, Forward Access Channel; FOCC-P, Forward Paging Channel; RECC-A, Reverse Access Channel.

the received SID with the stored SID of its home system, it determines whether it is "at home" or roaming and indicates this to the user. It also determines the first and last paging and access channels of the system.

The mobile then scans the "A" paging channels, tunes to the strongest one, again reads an overhead message (b), and completes its initialization by storing all parameters in the received message.

12.3.2 Registration

When the mobile has initialized, it makes the system aware of its presence. It scans the access channels of the system, tunes to the strongest one, and then sends a Registration message (c). The message includes the MIN, MSN, and other information about the mobile. The MSC checks the validity of MIN and MSN. When MIN and MSN are valid and do not indicate a problem (e.g., a MSN marked as a stolen mobile station), the MSC returns a Registration Confirmation message (d) on the associated forward channel. At this point, the mobile and MSC have sufficient information to handle calls from and to the mobile.

The mobile now enters the "idle" state, tunes to the strongest paging channel, and keeps reading overhead messages (e), updating its variable-memory when a parameter in an overhead message changes. The idle state ends when the mobile user originates a call or is called.

12.3.3 Originating Call

The user originates a call by keying the called number and then depressing the "send" button of the mobile station—Fig.12.3-2. The mobile then tunes to the strongest access channel and sends an Origination message (a). The message includes

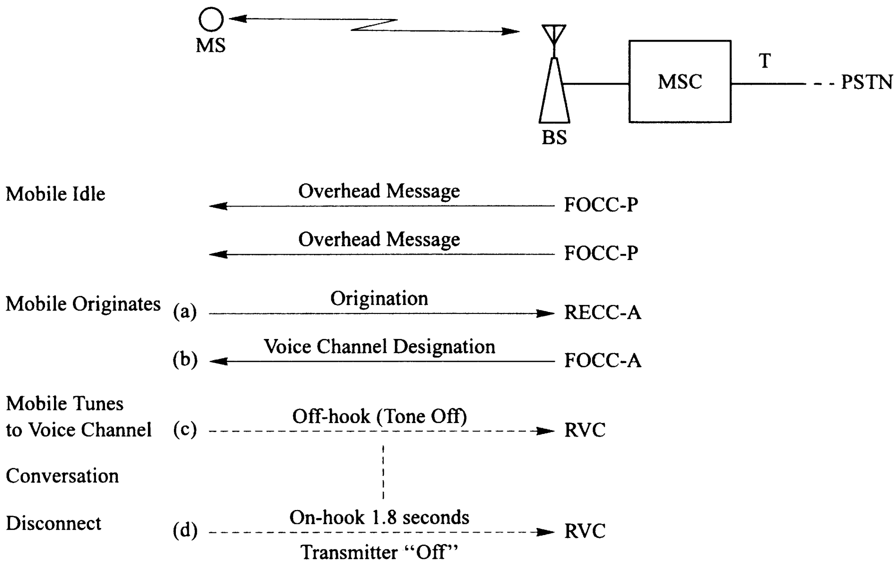


Figure 12.3-2. Call originated by mobile. FOCC-A, Forward Access Channel; FOCC-P, Forward Paging Channel; RECC-A, Reverse Access Channel; RVC, Reverse Voice Channel.

MIN, MSN, and the called number. The MSC then seizes an available trunk T to an exchange in the PSTN exchange, and an available trunk and associated voice channel in the BS that received the origination. It then signals the called party number on trunk T and sends an Initial Voice Channel Designation message in which the channel and attenuation code are specified by parameters CHAN and VMAC (b). The MSC also turns on the SAT on the voice channel (not shown).

The mobile then tunes its transmitter and receiver to the voice channel, sets its output power, returns the received SAT (not shown), and indicates off-hook (c). When the setup of the call reaches the called exchange, the mobile user hears ringing-tone (or busy-tone). Assuming that the called party answers, the conversation begins.

In this example, the mobile user disconnects first, by depressing the "end" button. The mobile sends an on-hook pulse (signaling tone "on") for 1.8 s and then turns off its transmitter (d). The MSC recognizes the disconnect request, turns off the BS transmitter of the voice channel, and releases the voice channel and associated BS trunk, and the trunk to the PSTN.

12.3.4 Terminating Call

Idle mobiles are tuned to the strongest paging channel (Fig. 12.3-3). In addition to the overhead messages (a), the forward paging channels FOCC-P carry Page messages, which inform the mobiles about incoming calls (b). These messages include the MIN of the called mobile. When a mobile reads a Page message and recognizes its own MIN, it determines the strongest access channel and sends a Page Response (c).

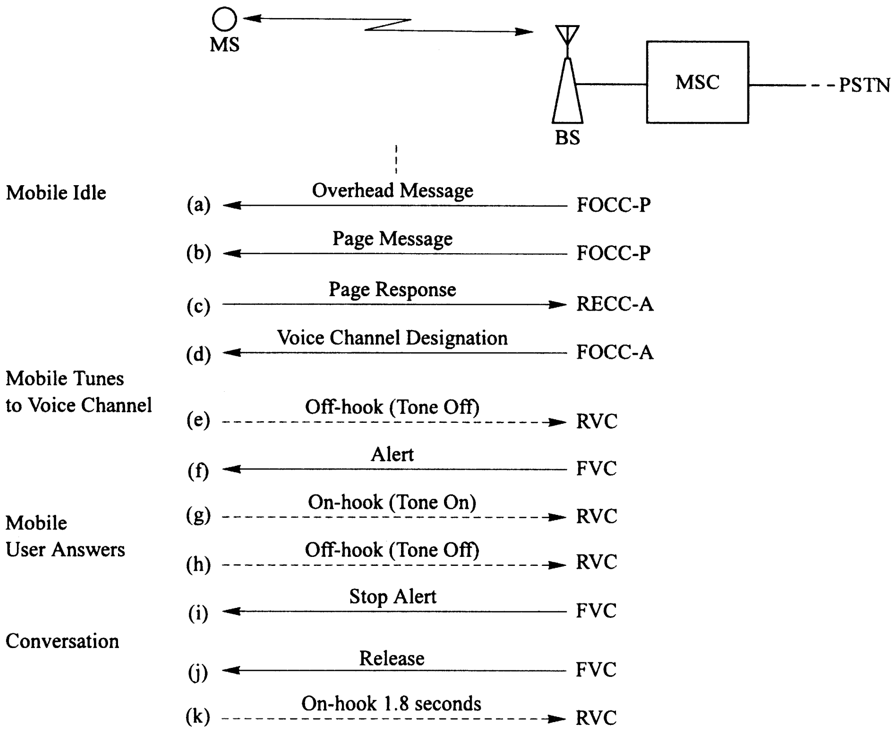


Figure 12.3-3. Call to mobile. FOCC-A, Forward Access Channel; FOCC-P, Forward Paging Channel; RECC-A, Reverse Access Channel; RVC, Reverse Voice Channel; FVC, Forward Voice Channel.

The MSC then seizes an available voice channel in the BS where the page response was received, starts to transmit SAT (not shown), and sends an Initial Voice Channel Designation message (d) on the access channel. Message parameters CHAN and VMAC specify the channel and the transmit power. The mobile tunes to the channel, sets its transmit power, and returns SAT and off-hook (supervision tone off) on the RVC (e). The MSC then sends an Alert message (f). In response, the mobile generates ringing-tone for its user and changes its state to on-hook (g). When the user answers, the mobile changes back to off-hook (h). In response, the MSC sends a Stop Alert message (i) and connects the BS trunk to the selected voice channel. The conversation can begin.

Assuming that the calling party disconnects first, the MSC sends a Release message (j), and the mobile acknowledges with a disconnect signal (going on-hook for 1.8 s) and then turns off its transmitter (k). Finally, the MSC releases the voice channel and associated BS trunk, and the trunk to the PSTN exchange.

12.3.5 Power Change and Handoff

When on a call, a mobile can receive a command on its voice channel to change its transmitter power or to tune to a different voice channel—see Fig. 12.3-4.

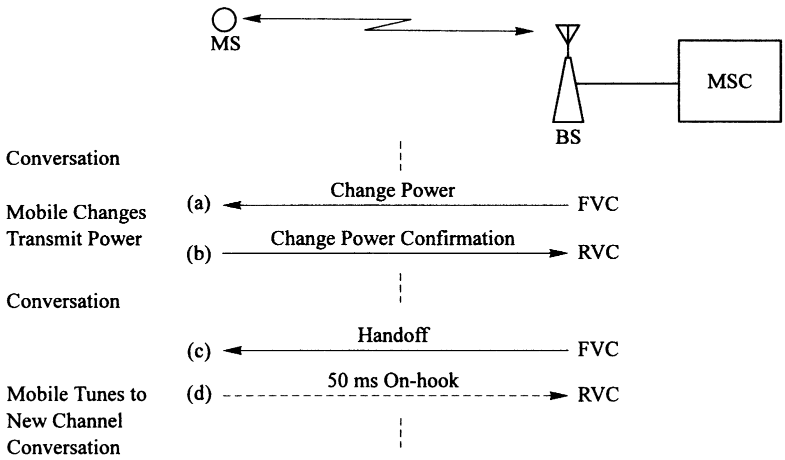


Figure 12.3-4. Power change and handoff. RVC, Reverse Voice Channel; FVC, Forward Voice Channel.

Power Change. A MSC monitors the mobile signal strengths, received by the base stations on their reverse voice channels. When it decides that a mobile should change its transmit power, it sends a "change power" Order (a) that includes a new VMAC value. The mobile then adjusts its transmit power and returns a "change power" confirmation message (b).

Handoff. In a handoff (see Section 12.1.7), the MSC sends a Handoff message (c). The message includes the new channel number CHAN and attenuation code VMAC. The mobile acknowledges with a 50-ms on-hook signal (d) and turns off its transmitter. It then tunes to the new channel, sets the new power level, and turns the transmitter on again.

12.4 AMPS MESSAGE FORMATS AND PARAMETERS

AMPS distinguishes two groups of messages, *overhead parameter messages* and *mobile control messages*. Overhead parameter messages are broadcast to all mobiles on the Forward Control Channels (FOCCs). Mobile control messages contain call-control messages for individual mobiles and—depending on the message—are sent on forward or backward control, or voice channels (FOCC, RECC, FVC, and RVC).

Before exploring the individual messages, two general aspects of AMPS message words need to be mentioned—see Fig. 12.4-1. In the first place, each word has a 12-bit P (parity) field, which is used for error control (Section 12.2.2). In the second place, the developers of AMPS, anticipating future additions to signaling in mobile systems, included a number of *reserved* (RSVD) fields in message words. In AMPS, a mobile or MSC sets the RSVD fields in its outgoing messages

T1T2 = 11	DCC	SID1	RSVD = 000	NAWC	OHD = 110	P	PW1
2	2	14	3	4	3	12	

T1T2 = 11	DCC	S	E	REGH	REGR	DTX	N-1	RCF	CPA	CMAX-1	END	OHD = 111	P	PW2
2	2	1	1	1	1	2	5	1	1	7	1	3	12	

T1T2 = 11	DCC	ACT = 0010	REGINCR	RSVD = 0000	END	OHD = 100	P	GAW2
2	2	4	12	4	1	3	12	

T1T2 = 11	DCC	ACT = 1001	BIS	RSVD = 00......0	END	OHD = 100	P	GAW9
2	2	4	1	15	1	3	12	

T1T2 = 11	DCC	ACT = 1010	MAXBUSY - PGR	MAXSZTR - PGR	MAXBUSY - OTHER	MAXSZTR - OTHER	END	OHD = 100	P	GAW10
2	2	4	4	4	4	4	1	3	12	

T1T2 = 11	DCC	REGID	END	OHD = 000	P	RIDW
2	2	20	1	3	12	

T1T2 = 11	DCC	= 010111	CMAC	RSVD = 00	= 11	RSVD = 00	= 1	WFOM	= 1111	END	OHD = 001	P	CFW
2	2	6	3	2	2	2	1	1	4	1	3	12	

Figure 12.4-1. Overhead Parameter message words. (From EIA/TIA 553. Reproduced with permission of TIA.)

to 00, . . . , 0 and does not examine these fields in received messages. We shall see later how these fields are used in post-AMPS mobile systems in the United States.

12.4.1 Overhead Parameter Message Format

An AMPS Overhead Parameter message (OPM) consists of several words. Figure 12.4-1 shows the most important OPM words [6]. All messages include parameter words PW1 and PW2. The other words are included only when necessary.

In all OPM words, the *type* field (T1T2) is set to 11. This distinguishes these words from words of mobile control messages on FOCC. The overhead (OHD) field identifies the word type:

OPM Word Type	Acronym	OHD
First parameter word	PW1	110
Second parameter word	PW2	111

Global action words	GAW	100
Registration ID word	RIDW	000
Control filler word	CFW	001

The global action words contain system parameters that were added at a later stage of AMPS development. The various GAWs are distinguishable by the codes in their action (ACT) fields. We limit the discussion to the GAWs shown in Fig. 12.4-l.

In word PW1, the NAWC (number of additional words coming) field indicates the number of subsequent message words. The latter words have END indicators that are set to 0 in all words, except the last one.

The control filler word (CFW) also has T1T2 = 11 but is not part of an overhead message. CFWs are sent when no other messages have to be transmitted. A CFW is similar to the fill-in signal unit (FISU) of Signaling System No. 7 (Section 7.3) but—unlike the FISU—also holds a number of parameters.

12.4.2 Overhead Message Parameters

This section describes the parameters shown in Fig. 12.4-1. They are denoted by acronyms and are listed in alphabetical order. At this point it is suggested that the descriptions are briefly perused and referred to when reading the sections that deal with signaling procedures.

BIS: Busy-Idle Status Indicator (in word GAW 9, which is the GAW with ACT = 1001). If BIS = 1, a mobile sending a message on a reverse control channel must check for an idle-to-busy transition on the associated forward control channel.

CMAC: Control-Channel Mobile Attenuation Code (in CFW). This indicates the mobile transmit power level to be used on reverse control channels.

CMAX-1 (in PW2). This is the number of access channels in a system, minus 1 (in a system with 10 access channels, CMAX − 1 = 9).

CPA: Combined Paging and Access Indicator (in PW2). When CPA = 1, the control channels in the system are combined paging and access channels. When CPA = 0, the system has separate paging and access channels.

DCC: Digital Color Code (in all OPM words). This is the color code of the transmitting base station (values: 00, 01, or 10; see Section 12.1.4).

DTX: Discontinuous Transmission Indicator (in PW2). This indicates whether a mobile with discontinuous transmission (Section 12.1.6), and transmitting on a voice channel, is allowed to reduce its transmit power when the user is silent:

DTX = 00	Power
DTX = 10	Limited (8-dB) power reduction allowed
DTX = 11	Any power reduction allowed

When the mobile user speaks, or a message is sent, the transmitter should return immediately to its normal power level.

E: Extended Address Indicator (in PW2). If E = 1, all mobiles have to include their complete mobile identification number MIN in messages sent on a reverse control channel. If E = 0, only roaming mobiles need to include the complete MIN. Resident mobiles send a partial MIN (without area code).

Maxbusy-PGR, Maxbusy-Other, Maxsztr-PGR, and Maxsztr-Other (in GAW 10). When a mobile needs to send a message on a reverse control channel, these parameters indicate the maximum number of times the mobile can try to seize a reverse control channel.

N − 1 (in PW2). This is the number of paging channels in the system minus one (in a system with five paging channels, $N - 1 = 4$).

RCF: Read Control Filler Indicator (in PW2). If set to 1, the mobile must read and copy the parameters in a control filter word (CFW) before accessing a reverse control channel.

REGH and REGR (in PW2). These are home and roamer registration indicators. When set to 1, the system allows registrations by resident and roaming mobiles, respectively.

REGID and REGINCR (in, respectively, RIDW and GAW 2). These are the registration identification and registration increment, which are integers used by mobiles for autonomous registration.

S: Serial Number Indicator (in PW2). If S = 1, mobiles have to include their MSN (mobile serial number) when sending a message on a reverse control channel.

SID1: System Identification, Part 1 (in PW1). The leading 14 bits of the 15-bit *System Identification* (SID). The least significant bit is omitted because, prior to sending a message, the mobile has decided to communicate with the "A" or "B" system and therefore knows whether SID is odd or even (Section 12.3).

WFOM: Wait-for-Overhead-Message Indicator (in CFW). If set to 1, mobiles have to await an overhead message before seizing a reverse control channel.

12.4.3 Mobile Control Messages

We now examine the most important mobile control messages. Some of these have already been mentioned in Section 12.3. Mobile control messages on the Forward and Reverse Control Channels (FOCC, RECC) are used for signaling between the mobile system and a mobile that has been turned "on" and is not involved in a call. When a mobile is on a call, the signaling is on the forward and reverse part of the voice channel (FVC, RVC) that has been allocated to the call.

12.4.4 Mobile Control Messages on FOCC

The messages shown in Fig. 12.4-2 consist of word 1 ($T1T2 = 01$) and word 2 or word 2* ($T1T2 = 10$).

Page Message. This alerts the mobile about an incoming call.

Registration Confirmation Message. This confirms the registration of a mobile.

Release Message. This is sent when the system rejects the registration of a mobile.

Reorder Message. This is sent when the system cannot process an origination, for example, when no voice channel is available.

Initial Voice Channel Designation Message. This specifies the initial voice channel assigned to the call, the color code of the channel, and the mobile attenuation code.

Word 1

T1T2 = 01	DCC	MIN1	P
2	2	24	12

Word 2

T1T2 = 10	SCC = 11	MIN2	RSVD = 0 - - - 0	ORDQ	ORD	P
2	2	10	6	3	5	12

Word 2*

T1T2 = 10	SCC ≠ 11	MIN2	VMAC	CHAN	P
2	2	10	3	11	12

Message	Words
Page	1, 2
Registration Confirmation	1, 2
Release	1, 2
Reorder	1, 2
Initial Voice Channel Designation	1, 2*

Figure 12.4-2. Mobile control messages on Forward Control Channel. (From EIA/TIA 553. Reproduced with permission of TIA.)

F = 1	NAWC	T	S	E	RSVD = 0	SCM	MIN1	P	Word 1
1	3	1	1	1	1	4	24	12	

F = 0	NAWC	RSVD = 0 - - - 0	ORDQ	ORD	LT	RSVD = 0 - - - 0	MIN2	P	2
1	3	5	3	5	1	8	10	12	

F = 0	NAWC	MSN	P	3
1	3	32	12	

F = 0	NAWC	DIGIT	DIGIT	DIGIT	DIGIT	DIGIT	DIGIT	DIGIT	DIGIT	P	4, 5...
1	3	4	4	4	4	4	4	4	4	12	

Message	Words
Registration	1, 2, 3
Page Response	1, 2, 3
Origination	1, 2, 3, 4...

Figure 12.4-3. Mobile control messages on Reverse Control Channel. (From EIA/TIA 553. Reproduced with permission of TIA.)

12.4.5 Mobile Control Messages on RECC

In these messages, indicator bit F is 1 in the first message word and 0 in succeeding words—Fig. 12.4-3. Messages always include words 1 and 2. Word 3 holds the MSN of the mobile and is included only if indicator S in the Overhead Parameter message is set to 1 (Section 12.4.2). Word 4 (and, when necessary, words 5, 6, etc.) appears in originations only and holds up to eight digits of the called number.

Page Response. This is sent by a mobile to acknowledge a page message.

Registration. This is sent by a mobile to announce its presence to the system.

Origination. This is sent by a mobile to request the setup of a call.

12.4.6 Mobile Control Messages on FVC

The type field in the messages is set to T1T2 = 10 (Fig. 12.4-4). All message types except "handoff" consist of word 1.

Alert. This is a request by the system to generate ringing-tone, to alert the user about an incoming call.

Stop Alert. This is a request to turn off ringing-tone.

Release. This orders the mobile to release the voice channel.

Change Power. This orders the mobile to change its transmit power.

T1T2 = 10	SCC = 11	PSCC	RSVD = 0 - - - 0	ORDQ	ORD	P	Word
2	2	2	14	3	5	12	1

T1T2 = 10	SCC ≠ 11	PSCC	RSVD = 0 - - - 0	VMAC	CHAN	P	Word
2	2	2	8	3	11	12	1*

Message	Word
Alert	1
Stop Alert	1
Release	1
Change Power	1
Send Called Address	1
Handoff	1*

Figure 12.4-4. Mobile control messages on Forward Voice Channel. (From EIA/TIA 553. Reproduced with permission of TIA.)

Send Called Address. This orders the mobile to send digits: for example, a called party number or a service access code (Section 3.5.1).

Handoff. This alerts the mobile that it is being handed off to a new cell and specifies the new voice channel, its color code, and the new transmit power. This message consists of word 1*.

12.4.7 Mobile Control Messages on RVC

$F = 1$ in the first message word, and $F = 0$ in succeeding words (Fig. 12.4-5).

Change Power Confirmation (word 1). The mobile confirms the receipt of a change power order.

F = 1	NAWC	T	RSVD = 0 - - - 0	ORDQ	ORD	RSVD = 0 - - - 0	P	Word
1	2	1	5	3	5	19	12	1

F = 0	NAWC	T	DIGIT	DIGIT	DIGIT	DIGIT	DIGIT	DIGIT	DIGIT	P	Word
1	2	1	4	4	4	4	4	4	4	12	2, 3...

Message	Words
Change Power Confirmation	1
Called Address	1, 2, ...

Figure 12.4-5. Mobile control messages on Reverse Voice Channel. (From EIA/TIA 553. Reproduced with permission of TIA.)

Called Address (word 1, word 2, and succeeding words when necessary). This is sent by a mobile and represents a called party number (for three-way calling) or a service access code.

12.4.8 Mobile Control Message Parameters

The parameters in mobile control messages are described below in alphabetical order of their acronyms [6].

CHAN: Channel. This identifies an initial or new voice channel (Figs. 12.4-2 and 12.4-4).

DCC: Digital Color Code. This is a binary number that identifies the color code (see Section 12.1.4) of the FOCC in overhead parameter and mobile control messages (Figs. 12.4-1 and 12.4-2).

DIGIT. This is a digit of the called address or service access code (Figs. 12.4-3 and 12.4-5).

E: Extended Address Indicator. $E = 1/0$ in a RECC message indicates that the message includes/does not include a word 2, which holds the area code (MIN2) of the mobile identification number (Fig. 12.4-3).

LT: Last Try Indicator. $LT = 1$ indicates a final attempt by a mobile to send a message on RECC (Fig. 12.4-3).

MIN1: Mobile Identification Number 1. This is the seven-digit directory number (exchange code + line number) of the mobile identification number (MIN), coded as a 24-bit binary number (Figs. 12.4-2 and 12.4-3).

MIN2: Mobile Identification Number 2. This is the three-digit area code of the MIN, coded as a 10-bit binary number (Figs. 12.4-2 and 12.4-3).

MSN: Mobile Serial Number. This is a number assigned by the manufacturer and stored in nonalterable memory of the mobile. It is included in RECC messages if requested by the system (indicator bit $S = 1$ in Overhead Parameter messages)—see Fig. 12.4.3.

NAWC: Number of Additional Words Coming. This indicates the number of subsequent words in the message.

ORD: Order Code. This identifies the type of an order or confirmation message. Most mobile control messages contain an ORD (Figs. 12.4-2 through 12.4-5). In these messages, $SCC = 11$ (a nonexistent color code). The Initial Channel Designation and Handoff messages do not include an ORD, and SCC represents the color code of the initial or new voice channel (00, 01, 10—see Section 12.1.4). The order codes for the messages described in the previous section are listed in Table 12.4-1.

TABLE 12.4-1 Color Codes (SCC) and Order Codes (ORD) in AMPS Mobile Control Messages

		On Channel Type:				
SCC	ORD	FOCC	RECC	FVC	RVC	Message
11	00000	x				Page message
11	00000		x			Origination, Page Response
11	00001			x		Alert
11	00011	x		x		Release
11	00100	x				Reorder
11	00110			x		Stop Alert
11	01000			x		Send Called Address
11	01000				x	Called Address Messages
11	01011			x		Change Power Order
11	01011				x	Change Power Confirmation
11	01101	x				Registration Confirmation
11	01101		x			Registration
00, 01, 10	—	x				Initial Voice Channel Designation
00, 01, 10	—			x		Handoff message

Source: EIA/TIA 553. Reproduced with permission of TIA.

ORDQ: Order Qualifier. This field is included in all messages that have an ORD. In most messages, ORDQ is not used and is coded 000. In the change power order and its confirmation, ORDQ holds the mobile attenuation code that specifics the transmit power (Section 12.1.6). In Registration messages, it differentiates autonomous and nonautonomous registrations.

PSCC: Present SCC. This is the color code of the present voice channel in FVC messages (Fig. 12.4-4). If PSCC does not agree with the SCC code specified in a channel-assignment or Handoff message, the mobile ignores the message.

S: S Indicator. This is included in RECC messages. When S = 1, the message includes a word 3 (Fig. 12.4-3).

SCM: Station Class Mark. This provides information on the transmitter of the mobile station (power class and capability for discontinuous transmission—see Section 12.1.6).

SCC: SAT Color Code. This appears in messages on FVC (Fig. 12.4-4). The values SCC = 00, 01, or 10 indicate an initial voice channel designation or a Handoff message and represent the color code of the designated voice channel. SCC = 11 indicates messages of other types, specified by ORD.

T: T Indicator. This indicator differentiates orders and confirmations in reverse messages. In RECC messages, T = 1 and 0 indicate, respectively, Orders and Order Confirmations (Fig. 12.4-3). In RVC messages, T = 1 and 0 indicate, respectively, an Order Confirmation and a Called Address message (Fig. 12.4-5).

VMAC: Voice Mobile Attenuation Code. This specifies the mobile transmit power on the voice channel (Figs. 12.4-2 and 12.4-4).

Coded DCC. All messages on FOCC include the parameter DCC that indicates the color code of the transmitting base station. A mobile sending a message on a RECC transmits a seven-bit coded DCC that represents the received DCC:

Received DCC	Coded DCC
00	000 0000
01	001 1111
10	110 0011

In contrast with the other parameters described here, the coded DCC is not contained in a message word, but follows the dotting (DOT) and word synchronization (WS) sequences—see Fig. 12.2-1(*c*).

12.4.9 Color Codes and Message Acceptance

Message Acceptance on Control Channels. The DCC and coded DCC on control channels are the digital counterparts of the SATs on voice channels (Section 12.1.4) and have the same purpose. A mobile discards a received mobile control message in which DCC is different from the DCC being received in overhead messages (Figs. 12.4-1 and 12.4-2). Likewise, a base station ignores a message received with a coded DCC that does not represent its DCC (Fig. 12.4-3), because it must have been sent by a mobile that is signaling to another BS.

Message Acceptance on Voice Channels. The Initial Voice Channel Designation and Handoff messages include the color code of the allocated channel. A mobile discards a FVC message in which PSCC does not have the expected value (Fig. 12.4-4). There is no comparable color-code checking for messages on RVC.

12.5 AMPS SIGNALING PROCEDURES

This section discusses a number of AMPS signaling procedures, adding details that were omitted in Section 12.3. While reading this material, it is helpful to look up the messages in Figs. 12.4-1 through 12.4-5, and the parameter descriptions in Sections 12.4.2 (overhead message parameters) and 12.4.8 (mobile control message parameters).

12.5.1 Mobile Initialization

The general procedure has been described in Section 12.3.1. We focus here on the retrieval by the MS of the system-specific data from overhead parameter messages.

Step 1. When the MS is turned "on," it scans the dedicated Forward Control Channels (FOCCs) of the "A" system (channels 333–313), if its home system is an “A” system, or the “B” system (channels 334–354), if its home system is a “B” system. It tunes its receiver to the strongest one. After receipt of the first overhead message (Fig. 12.4-1), the MS determines the first and last paging channels of the system from the value of $N-1$,

	“A” Systems	“B” Systems
First paging channel	333	334
Last paging channel	$(333-N+1)$	$(334+N-1)$

and stores this information in its temporary memory.

Step 2. The mobile then tunes to the strongest paging channel, waits for another overhead message, and compares the received system identifier (SID1) with the SID of its “home” system, which is stored in its semipermanent memory. If the identifiers match, the MS is “at home.” If not, the MS informs its user by turning “on” its “roaming” light.

Step 3. The MS determines the first and last access channels from parameters CPA and CMAX-1. If CPA = 1, the system has combined access and paging channels, and the first and last access channels are the same as the first and last paging channels. If CPA = 0, the system has separate access and paging channels, and the first and last access channels are:

	“A” systems	“B” systems
First channel	$333-N$	$334+N$
Last channel	$(333-N-\text{CMAX}+1)$	$(334+N+\text{CMAX}-1)$

The MS then stores the first and last access channels in its temporary memory.

Step 4. The MS awaits another overhead parameter message and stores the other overhead message parameters and indicators (Section 12.3.2).

The MS is now initialized and in the “idle” state. It remains tuned to the strongest paging channel. While in this state it keeps monitoring the overhead messages and updates its memory when an overhead parameter changes. The idle state changes when the MS user originates, and when the MS receives a paging message, indicating a terminating call.

If the initialization procedure fails, the MS tries to contact the other system (the “B” system if the mobile homes on an “A” system, and vice versa).

12.5.2 Seizing a Reverse Access Channel

When an idle MS needs to access the system (send a message), it has to seize a Reverse Access Channel. This is done in the following series of steps.

Step 1. The mobile scans the access channels, locks on to the strongest one, reads an overhead message, and examines parameter RCF. If RCF = 0, the mobile sets its transmitter to maximum power and starts step 3. If RCF = 1, it goes to step 2.

Step 2. The MS reads a control filler word CFW (Fig. 12.4-1), sets its transmit power to the value in CMAC, and, if WFOM = 1, also reads another overhead message, updates its parameters, and then goes to step 3.

Step 3. Any idle MS in a cell can seize an access channel. In order to minimize "collisions" (simultaneous seizure of an access channel by more than one mobile), the mobile first examines the BI (busy idle) bits on the Forward Access Channel, which indicate whether the associated reverse channel is idle (Section 12.2.2).

If the channel is busy, the mobile waits for a random time (0–200 ms) and repeats this step. Up to NBUSY-PGR or NBUSY-OTHER busy occurrences are allowed for, respectively, page response and other messages. If the channel is idle, the MS goes to step 4.

Step 4. The MS seizes the channel and starts its transmission. In systems with BIS = 0, the mobile transmits the entire message. If BIS = 1, the system has a second defense against "collisions." The mobile then has to keep monitoring the BI bits on FOCC. BI turning to busy before 56 message bits have been sent indicates a collision, and BI not changing to busy after 104 bits have been sent indicates that the message is not being received. In either case, the mobile stops transmitting. The number of allowed RECC seizures is MAXSZTR-PGR or MAXSZTR-OTHER. As long as this limit is not exceeded, the mobile waits for a random time and returns to step 3.

12.5.3 Registration

A mobile always tries to register when it has completed its initialization (Section 12.3.2). Before sending a Registration message, the resident or roaming MS first examines REGH or REGR and attempts to seize an access channel only when allowed by the system. The mobile system acknowledges a received registration with a Registration Confirmation message (Fig. 12.3-1).

12.5.4 Determining the MS Location

The service area of a MSC is divided into a number of *location areas*, which are clusters of adjacent cells (Section 12.1.1). When accessing the system with a Registration or Origination message, the MS transmits on the strongest access channel (i.e., the channel of the nearest base station). The MSC derives the current location area of the MS from the identity of the BS that has received the message and enters it into its record for the MS.

Autonomous Registration. Mobiles with "autonomous registration" capability also reregister periodically, on command from the MSC. Autonomous registration is governed by overhead parameters REGID and REGINCR, which are copied by the mobiles. The MSC increases REGID after a certain number of received registrations, by including a RIDW word with the new REGID value in its overhead parameter messages (Fig. 12.4-1).

On each registration, the MS copies the current values of REGID and REGINCR, say, 34567 and 500, and then calculates the value of an internal parameter: NXTREG = 500 + 34567 = 35067. When the system has increased REGID to a value that exceeds NXTREG, the mobile reregisters, determines a new NXTREG value, and so on. On receipt of each reregistration, the MSC determines and stores the new location area, thus keeping track of the MS location.

12.5.5 Paging the MS

Consider a MSC area that contains N cells. If a MSC receives a call for a MS that has registered in its service area, it could transmit a page message on all N paging channels (FOCC-P). However, this is very ineffective if N is large (which is the case in metropolitan MSC areas), because only one of the N messages is received by the MS. The large number of ineffective page messages tends to overload the FOCC-P channels.

This is the reason for autonomous MS registration, which enables the MSC to keep track of the location areas of the individual MSs in its service area. A MSC starts by transmitting page messages on only the FOCC-Ps in the last known location area of the MS. If MS is still in this area, it responds with a Page Response message. If the MSC does not receive a page response within a few seconds, it assumes that the MS has moved to another location area and repeats the paging message, this time on all FOCC-Ps.

12.5.6 Supplementary Services

Most cellular systems in the United States can provide call waiting, call forwarding, and three-way calling to mobiles that have subscribed to these services [7]. These services are supported by the signals and messages described so far.

To activate or deactivate call forwarding, the mobile user sends an origination message in which the digits represent a service access code. All access codes start with * (asterisk). In this way, the MSC can distinguish feature activations/deactivations and originations.

Mobile users who are on a call and have call-waiting service are informed by a tone that another call has arrived. The user can then switch back and forth between the original and the new call, by sending "flash" signals (Section 12.2.1).

Mobile users who are on a call and have three-way calling service can initiate a call to a third party by sending a flash. The MSC responds with a "send called address" message, the mobile then sends the "called address" message, and the second call is added to the connection.

12.5.7 Protection Against Cloning

A mobile identifies itself to a MSC by including its mobile identification number (MIN) and serial number (MSN) in its registration, origination, and page-response messages, and the MSC serves the mobile only when it has verified these parameters. This gives some protection against customer fraud. For example, databases in cellular systems maintain lists with the MSNs of stolen mobile stations and do not give service to these stations. However, this does not protect against persons with the required equipment and technical knowledge who pick up messages on a reverse control channel, and thereby acquire valid combinations of MIN and MSN. These people are also capable of entering one of these combinations in a stolen MS. This activity is known as "cloning."

Mobile systems accept originations from these mobiles and charge the calls to the owner of the MS whose MIN-MSN combination has been cloned. The fraud is detected when this owner complains about charges for unknown calls on his monthly invoice, but this happens on average one half-month after the illegally altered MS has started making calls. In the United States, the annual losses resulting from cloning are estimated at about $500,000,000.

To protect against cloning, the operators of some mobile systems are issuing four-digit personal identification numbers (PINs) to MS users. When the user originates or responds to a page message, the MSC allocates a voice channel and sends an initial voice channel designation message. It then sends a short tone burst to the mobile, which prompts the user to send the PIN number. Only after the PIN has been validated does the MSC cut through the connection between the trunk to the PSTN and the BST and its associated voice channel (Fig. 12.1-3).

The protection offered by this arrangement is not complete. It is possible to construct equipment that can tune to a Reverse Access Channel (RECC-A) to capture a MIN + MSN combination, then tune to the associated Forward Access Channel (FOCC-A) to obtain the identity of assigned voice channel (VC), and finally tune to RVC and capture the PIN code.

A more powerful anticloning arrangement is described in Section 12.6.7.

12.6 SIGNALING IN IS-54 CELLULAR SYSTEMS

Beginning in the early 1990s, a number of second-generation cellular systems were deployed in the United States and abroad [8–11].

Two second-generation ion mobile cellular systems have been defined for use in the United States. These systems are known as the IS-54 and IS-95 systems—their names refer to the EIA/TIA Interim Standards in which they have been specified [12,13]. Both have digital multiplex voice channels. IS-54 uses *time-division multiple access* (TDMA), and IS-95 is based on *code-division multiple access* (CDMA). This section discusses the signaling in IS-54. IS-95 is discussed in the next chapter.

12.6.1 Dual-Mode Systems

Because of the large installed base of AMPS mobiles and AMPS cellular networks, IS-54 and IS-95 are *dual-mode* systems. The networks are implemented as additions to—or partial replacements of—existing AMPS networks and include both digital and analog (AMPS) RF channels. Likewise, the IS-54 and IS-95 mobile stations are dual-mode and can operate with analog and digital channels. In addition, IS-54 and IS-95 networks and mobiles are capable of using their respective signaling protocols and the AMPS protocol. This means that any MS can be served by any network. For example, an IS-54 mobile operating in an IS-95 network uses AMPS signaling and, if the MS originates a call, the network assigns an AMPS voice channel.

12.6.2 The IS-54 System

The digital channels of IS-54 use *differential quaternary phase-shift keying* (DQPSK), in which the phase of a signal with frequency f_c is changed 24,300 times per second. Each phase change has one of four values that represent the values of two consecutive bits. The bit rate in the channel is thus 48.6 kb/s. The power spectrum of the digital signal fits within 30 kHz, which is also the bandwidth of AMPS channels.

TDMA Traffic Channels. The 48.6-kb/s bit stream on the digital channels is divided into frames of 1994 bits (Fig. 12.6-1). The frame duration is 40 ms, giving a repetition rate of 25 frames/s.

Each frame is time-divided into six 6.67-ms time slots. Communications between the mobiles and a base station occur in TDMA bursts, during predetermined time slots of each frame. At present, the TDMA traffic channels are "full-rate," using two time slots in each frame (slots 1 and 4, or 2 and 5, or 3 and 6). A digital channel thus accommodates three full-rate TDMA traffic channels. "Half-rate" traffic channels, using only one time slot, are under development.

A time slot has 260 bits of digital speech (or other user information). The user information rate on full-rate channels is thus 13 kb/s. The A/D and D/A converters

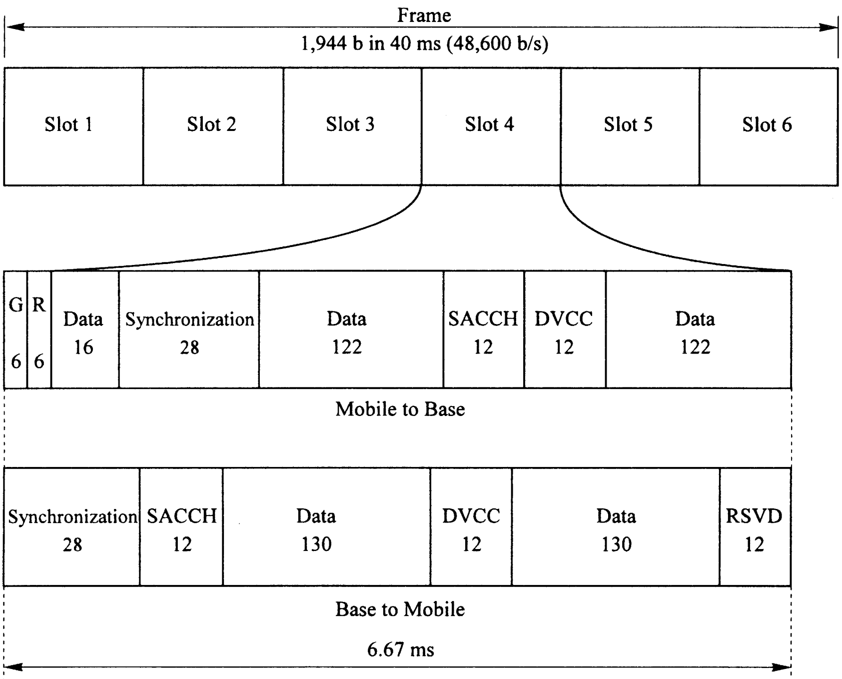


Figure 12.6-1. IS-54 frame structure. G, Guard time; R, ramp time; DVCC, digital verification color code; RSVD, reserved for future use. (From IEEE *Commun. Mag.* **29**(6). Copyright © 1991 IEEE.)

for speech in the mobiles and base stations operate at 8 kb/s. Before transmission, five forward error-correction bits are added to each eight-bit speech sample, yielding the bit rate of 13 kb/s.

Each time slot includes a specific 28-bit synchronization sequence. This enables the receiver in a mobile to lock on to its assigned time slots.

IS-54 has 255 *digital verification color codes* (DVCCs). They appear as 12-bit groups (eight information bits and four forward error correction bits) in each time slot. DVCC is the digital counterpart of the SAT tones on analog voice channels (Section 12.1.4). Mobiles and base stations ignore the received information in time slots with incorrect DVCCs.

System Capacity. Compared with AMPS, IS-54 considerably improves the RF system capacity, because the 30-kHz bandwidth required for one AMPS voice channel can also accommodate three IS-54 traffic channels. AMPS cells in seven-cell clusters have up to 56 voice channels and a traffic capacity of 42 erlangs (Section 12.1.5). Suppose that 28 AMPS voice channels in a cell are replaced by 3×28 digital traffic channels. This doubles the number of channels, and the traffic capacity, of the cell.

This is important in mobile systems that serve areas with high traffic densities (erlangs per square mile). In AMPS, this would require a large number of small cells, and this has two disadvantages. In the first place, high traffic densities often occur in metropolitan areas, where suitable locations for base stations are hard to find. In the second place, in small cells, the probability that a call has to be handed off is rather high. Replacing a number of AMPS by digital traffic channels allows the use of fewer, and larger, cells.

12.6.3 IS-54 Dual-Mode Signaling

In IS-54 systems, the dedicated AMPS control channels (channels 313–333 and 334–354, in the "A" and "B" systems—see Section 12.1.2) are used by AMPS and IS-54 mobiles. Some high-traffic IS-54 systems may require additional control channels. Therefore, channels 688–708 (in "A" systems) and 737–757 (in "B" systems) have been designated as secondary dedicated control channels. These channels can be used by IS-54 mobiles only, because AMPS mobiles are not designed to use them.

Messages on Control Channels. The messages on IS-54 control channels must be compatible with AMPS and IS-54 mobiles, and all AMPS messages described in Section 12.4 are therefore included without change.

To accommodate the additional information in AMPS overhead parameter and mobile control messages for IS-54 mobiles, use is made of the RSVD (reserved) message fields, which are ignored by AMPS mobiles and networks.

Messages on Voice and Traffic Channels. The signaling on the analog voice channels remains as in AMPS Each IS-54 traffic channel has two associated control

(*a*)

T1T2 = 11	DCC	SID1	RSVD = 000	NAWC	OHD = 110	P
2	2	14	3	4	3	12

(*b*)

EP	AUTH	PCI
1	1	1

Figure 12.6-2. Word PW1 of overhead parameter message: (*a*) AMPS (see Fig. 12.4-1) and (*b*) use of RSVD field in IS-54. (From IS-54-B. Reproduced with permission of TIA.)

channels, which carry signaling messages during the times that the channel has been allocated to a mobile. The Slow Associated Control Channel (SACCH) uses 12 bits in each time slot—see Fig. 12.6-1. On SACCH, a signaling message word consists of 132 bits (66 for information; 66 for forward error correction). This gives a rate of $2 \times 12/2 = 12$ information bits per frame, or $12 \times 25 = 300$ information bits/s, which is slower than in AMPS. Most mobile control messages are sent on the SACCH.

The Fast Associated Control Channel (FACCH) messages are sent as short bursts in the user information field. This is comparable with blank-and-burst technique for messages in AMPS voice channels (sec Section 12.2.2). A FACCH message word consists of 65 information bits and 65 forward error-correction bits.

The FACCH is used primarily for handoff messages, where speed is important.

12.6.4 Determination of the System and Mobile Type

In the initial contact between a mobile and a mobile system, an IS-54 mobile needs to determine the type of the mobile network (AMPS or IS-54), and an IS-54 system needs to determine the type of the mobile.

An IS-54 system indicates its type—and other capabilities—in the three-bit RSVD field of word PW1 in the overhead parameter messages on FOCC—see Fig. 12.6-2. If *protocol control indicator* (PCI) = 1, the control channel belongs to an IS-54 system and can assign digital traffic channels.

An IS-54 mobile uses the RSVD fields of word 2, in the messages it sends on RECC—see Fig. 12.6-3. In these messages, an IS-54 mobile sets the *mobile protocol*

(*a*)

F = 0	NAWC	RSVD = 0 - - - 0	ORDQ	ORD	LT	RSVD = 0 - - - 0	MIN2	P
1	3	5	3	5	1	8	10	12

(*b*)

MSG TYP
5

EP	SCM(4)	MPCI	SDCC1	SDCC2
1	1	2	2	2

Figure 12.6-3. Mobile control messages on reverse control channel. (*a*) AMPS (see Fig. 12.4-3) and (*b*) use of RSVD fields in IS-54. (From IS-54-B. Reproduced with permission of TIA.)

capability indicator (MPCI) to 01. Messages with this MPCI can include words that are not used in AMPS.

12.6.5 Order Qualifiers

In AMPS call-control messages, the ORD field identifies an order (or the acknowledgment of an order), and the ORDQ field qualifies the order—see Figs. 12.4-2 through 12.4-5. In IS-54 messages on the FOCC (forward) and RECC (reverse) control channels, the order is qualified by the contents of ORDQ in combination with the contents of a five-bit *MSG TYP* field.

In FOCC messages, MSG TYP occupies five bits in the RSVD field of word 2 (Fig. 12.4-2). In RECC messages, MSG TYP is in the five-bit RSVD field of word 2 (Fig. 12.4-3).

12.6.6 Assignment of Digital Traffic Channels

When setting up a connection, the MSC needs to assign a channel of a type that is compatible with the types of available channels in the cell, and with the channel types that can be handled by the mobile.

When an IS-54 mobile originates a call, the ORDQ and MSG TYP fields in its origination message indicate, among other things, whether the mobile can handle analog voice channels, and/or full-rate, and/or half-rate digital traffic channels. The MSC uses this information to allocate a channel of the proper type.

When paging an IS-54 mobile, the MSC uses the ORDQ and MSG TYP fields in its paging message to indicate the types of available channels. The mobile includes the channel types it can handle in its page response message, and MSC then selects a channel of a compatible type.

To conclude this section, we examine a number of IS-54 features and the messages and parameters that support them.

12.6.7 Authentication

This feature is a defense against the cloning of mobiles (Section 12.5.7). It is based on the CAVE (cellular authentication and voice encryption) algorithm, which produces an 18-bit result AUTHR (authentication response), from the values of four inputs:

	Input	Acronym	Length (bits)
1	Random number generated by MSC	RAND	32
2a	In registrations and page responses: the first part of the mobile identification number	MIN1	24
2b	In originations: the last six digits of the called number (BCD coded)	6DIG	24
3	Mobile serial number	MSN	32
4	Shared secret data for authentication	SSD_A	64

In a MS with authentication capability, its SSD_A(M) is stored in semipermanent memory, along with MIN and so on, and the MS can execute CAVE.

Mobile networks have (or will have) *authentication centers* (AUCs). An AUC stores the parameters MSN(A) and SSD_A(A) associated with individual MINs of mobiles with authentication capability and can also execute CAVE.

In the authentication procedure, the MS and AUC execute CAVE. The MS uses its stored SSD_A(M), and the AUC uses the SSD_A(A) that is associated with the MIN of the mobile. The other inputs to the algorithm are the same at MS and AUC.

Suppose that the mobile and the AUC produce the results AUTHR(M) and AUTHR(A). The AUC receives a copy of AUTHR(M) and compares it with its result. If the AUTHRs match, AUC concludes that the mobile stores the proper SSD_A(M) value, namely, SSD_A(A), and must be "authentic." Only authentic mobiles will receive service from the network.

Authentication can take place when a mobile registers, originates a call, or responds to a page message. Figure 12.6-4 shows the procedure for originating mobiles. The message parameters that are significant for authentication are shown inside square brackets. The MSC indicates its authentication capability in its parameter overhead messages, by setting the authentication indicator in word PW1 to AUTH = 1 (see Fig. 12.6-2), and including two global action words each of

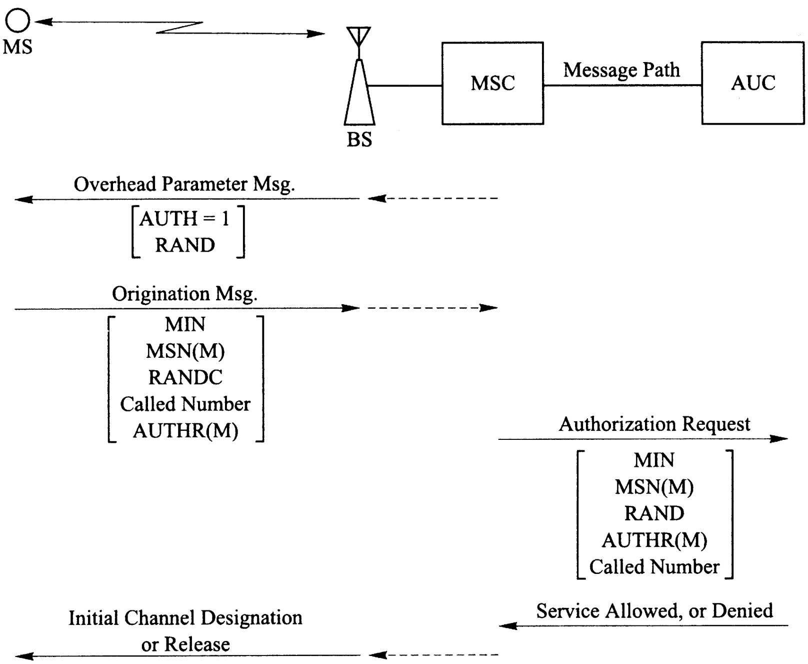


Figure 12.6-4. Authentication of originating mobile. (From TIA/EA TSB-51. Reproduced with permission of TIA.)

which holds 16 bits of RAND. The MSC frequently generates new RAND values, say, every minute.

The originating MS notes that AUTH = 1 and therefore executes CAVE, using the most recently received RAND, the final digits of the called number (6DIG), and its internally stored parameters MSN(M) and SSD_A(M). It then sends an origination message that includes, in addition to MIN, MSN(M), and the called number, a word that holds RANDC and AUTHR(M). RANDC contains the eight high-order bits of the RAND that was used in the calculation of AUTHR(M).

The MSC may have changed its RAND after MS has executed CAVE. Therefore, on receipt of the origination message, it compares the eight high-order bits of its current and previous RANDs with RANDC, to determine the RAND used by the MS.

The MSC then sends an "authentication request" message to the authentication center. AUC then executes CAVE, using MIN, MSN(M), RAND received in the message, and its stored SSD_(A). It then compares its result AUTHR(A) with AUTHR(M). If the AUTHRs match, AUC sends an "allow service" message to the MSC. If the AUTHRs do not match, a "deny service" response is sent. The signaling messages between MSC and AUC are discussed in Chapter 19.

When a MS registers, or sends a page response, it uses MIN1 instead of 6DIG for the execution of CAVE and does not include a called number in its message. If AUC receives an authorization request that does not include a called number, it also uses MIN1 for its CAVE.

It is next to impossible to defeat the authentication procedure, because SSD_A is not transmitted on RF channels. It is possible to pick up, on a reverse control channel, combinations of parameters sent by mobiles in their origination messages, but AUTHR(M) is of no use for cloning, because its value is different for each call of the mobile, since it depends on RAND and the called number.

12.6.8 Voice Privacy

This feature protects the mobile user against unauthorized "listening in" on his calls. This is done by scrambling the speech on digital traffic channels before it is fed into the RF transmitter, and unscrambling it at the receiving side of the channel. The voice privacy procedure is an extension of the CAVE procedure.

A mobile with voice privacy capability stores a second 64-bit "shared secret data" parameter SSD_B in its memory. SSD_B is also stored at the authentication center.

During the setup of a call, the MS and the AUC execute CAVE, using the inputs RAND, MIN1, MSN, and SSD_B, which produces two "voice privacy masks" (one for each direction of transmission). The AUC includes the masks in its response message to MSC.

A mask consists of 260 bits, each of which corresponds to a bit in the 260-bit user data fields shown in Fig. 12.6-1. Before transmitting a time slot, the MSC scrambles the user data fields that contain the digitally coded speech, by forming the "exclusive-or" of each data bit and its corresponding bit in the MSC-to-MS voice privacy mask.

The mobile does the same operation on the received data, using its copy of the mask. In the same way, before transmitting a time slot, the MS scrambles the data with the MS-to-MSC mask, and the MSC performs the same operation on the received data.

If the inputs to the algorithm at the MS match those at the MSC, the voice-privacy masks at MS and MSC match also, and the original data are restored. Both masks depend on SSD_B, which is not transmitted over the air. It is therefore practically impossible for outsiders to generate the masks.

12.6.9 Mobile-Assisted Handoff

The AMPS handoff procedure has been described in Sections 12.1.7 and 12.3.5. In this procedure, the signal strength of the reverse voice channel used by the mobile is measured—on command from the MSC—by the base station of the cell that currently serves the mobile and by a number of base stations in adjacent cells. As a result of the measurements, the MSC may decide to handoff the mobile to the cell that receives the strongest signal. This requires—in each cell—one or more measurement receivers that can be tuned to reverse channels that are not used in the cell itself.

In mobile-assisted handoff, an IS-54 mobile that is in conversation on a traffic channel can—on command from the MSC—make signal strength measurements on forward channels and report its results. This eliminates the need for measurement receivers in base stations and also reduces the handoff work at the MSC, which is important in high-traffic cellular systems. Measurements by an IS-54 mobile that is in conversation, and served by a traffic channel, are possible because the mobile receiver needs to be tuned to the assigned traffic channel only during two of the six time slots in a frame and can be used for measurements during the other time slots.

The signaling for a handoff measurement, which requires several new messages, is shown in Fig. 12.6-5. The MSC orders the MS to start measurements with a measurement order (a) that identifies up to 12 channels to be measured in cells adjacent to the cell that is serving the MS. The MS responds with an acknowledgment message and starts the measurements. The measurements continue until the MS receives a stop measurements order (c). The MS acknowledges the order (d) and stops its measurements. It then sends a channel-quality message (e) that includes the received signal strength indicators (RSSI) of the currently used channel and of the channels specified by the MSC. The value of RSSI ranges from 0 (very weak signal) to 31 (very strong signal). The MSC then decides whether to handoff the mobile to another cell. The measurement and handoff orders are sent on FACCH. The other messages can be sent on FACCH or SACCH.

IS-54 mobiles can be handed off to an analog or digital channel in an adjacent cell, regardless of the type of the current channel.

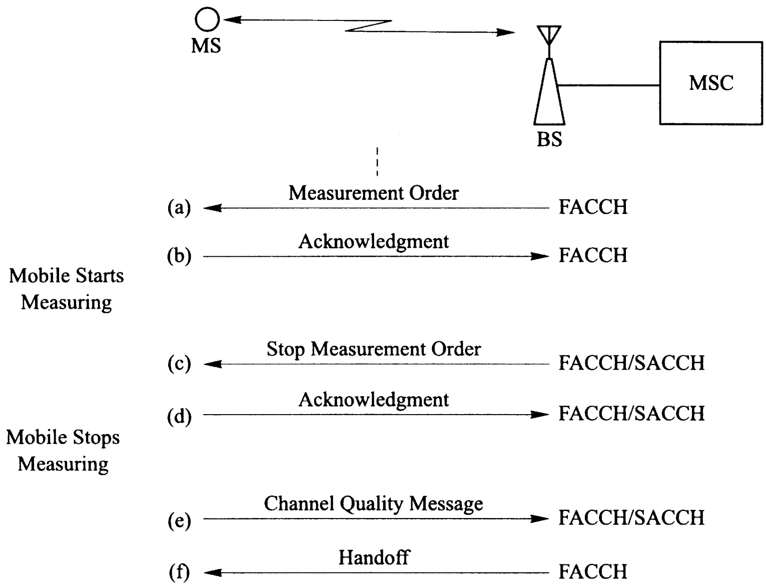


Figure 12.6-5. Mobile handoff signaling. FACCH, Fast Associated Control Channel; SACCH, Slow Associated Control Channel. (From IS-54-B. Reproduced with permission of TIA.)

12.7 INTRODUCTION TO THE GSM CELLULAR SYSTEM

The first-generation cellular systems developed in Europe were patterned after the AMPS system. However, each system has its own national standards, and a MS designed to work with the cellular system of a particular country cannot roam in another country.

In 1982, the Conférence Européenne des Postes et Télécommunications (CEPT) established a "Groupe Spéciale Mobile" (GSM) to start the definition of a Pan-European standard for a digital second-generation cellular system [8,9]. The resulting system is known today as GSM (Global System for Mobile Communications). In 1988, the responsibility for GSM standards was transferred to the European Telecommunications Standards Institute (ETSI).

GSM was first deployed in 1993 and has gained wide acceptance, in part because GSM mobiles can roam in any country with a GSM network. As of November 1994, it was serving two million subscribers in 26 European countries and had been adopted by another 26 countries outside Europe.

GSM radio channels are in the 900-MHz band. In the United Kingdom and Germany, this band was already allocated to other services, and these countries use the DCS1800 mobile system, which is very similar to GSM, but operates in the 1.8-GHz band.

GSM is a rather complicated system. This section presents a brief overview [8–11,14]. Signaling between a MS and a GSM network is discussed in Sections 12.8 and 12.9.

Most of the AMPS/IS-54 terminology also applies to GSM, but there are some differences. For example, in GSM documentation, cellular networks are known as public land mobile networks (PLMNs), base stations (BSs) are sometimes denoted as base station subsystems (BSSs), and *location updating* is the equivalent of MS registration.

12.7.1 GSM Overview

GSM is a single-mode system and has digital radio channels only. GSM mobile stations can receive service from GSM networks only.

GSM has several similarities with IS-54 (more correctly: IS-54, whose definition started in 1987, is similar in several respects to GSM). Both use a combination of frequency-division multiple access FDMA (RF carrier channels at different frequencies) and time-division multiple access TDMA (frames and time slots on the RF channels).

Like IS-54, GSM includes provisions for mobile-assisted handover (the European term for handoff), MS authentication, and *ciphering* (voice privacy). Moreover, it has been designed taking into account ISDN services, such as circuit- and packet-switched data communications.

12.7.2 GSM Radio Channels

A GSM network has 124 bidirectional RF channels, consisting of a *downlink* (forward BS to MS) channel and an *uplink* (reverse, MS to BS) channel [10,14]. The GSM radio channels are spaced at 200-kHz intervals. The uplink and downlink channels occupy the frequency bands of 890–915 and 935–960 MHz. The center frequency of an uplink channel is 45 MHz below that of the downlink channel. The bit rate on RF channels is 270 kb/s in each direction.

Each RF (physical) channel contains a number of time-division multiplexed *logical* channels. We distinguish *dedicated* and *common-control* channels:

Dedicated Channels. These are point-to-point (between a MS and a BSS) bidirectional channels. There are two groups of dedicated channels: Traffic Channels (TCHs) and Stand-Alone Dedicated Control Channels (SDCCHs). A Slow and a Fast Associated Control Channel (SACCH and FACCH) is associated with each TCH and SDCCH. The combination of a TCH and its SACCH and FACCH is known as a Traffic and Associated Control Channel (TACH).

A BSS has a pool of SDCCHs and a pool of TACHs. A SDCCH is allocated to a MS for call setup signaling and is released when this signaling is complete. When the call setup is successful, a TACH is allocated to the MS for the duration of the call. The TCH carries the user's speech or data, and the SACCH carries the "slow" signaling. When "fast" signaling is needed during a call (e.g., when the MS has to be handed over), the user communication is suppressed for a short time, and the TCH functions as a FACCH (like the "blank and burst" technique used in AMPS—see Section 12.2).

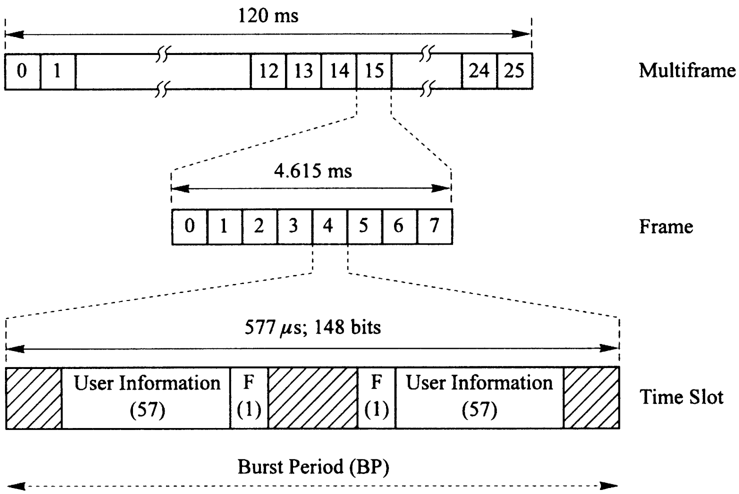


Figure 12.7-1. Frame format for TACH. The F bits indicate whether the time slot is a TCH or FACCH slot. (From *IEEE Commun. Mag.* **29**(6). Copyright © 1991 IEEE.)

Figure 12.7-1 shows the time slots, frames, and multiframes of TACHs. Each frame contains eight time slots, which have a duration of 0.557 ms and contain 148 bits, namely, 114 information bits, 2 flag bits (F), and 32 bits for RF transmission functions. A frame consists of eight time slots and has a duration of 4.615 ms. Twenty-six frames form a *multiframe* that has a duration of 120 ms.

A RF carrier can be used for eight TACHs, each of which uses a particular time slot. When a TACH has been allocated, the MS and the BS transmit in *burst periods* (BPs), which occur in the time slot of the TACH. The TCH is sent in 24 BPs of the multiframe (frames 0–11 and 13–24). The SACCH is sent in frame 12, and no BP is transmitted in frame 25.

The information bit rate on a TCH or FACCH is therefore $24 \times 114/(0.120) = 22.8$ kb/s. The GSM speech coder produces 13 kb/s, and forward error-correction of the speech (or user data) bits adds 9.8 kb/s. The information bit rate on a SACCH channel is $114/(0.120) = 0.95$ kb/s.

Eight SDCCHs share one time slot on a RF carrier. Their information bit rate is thus one-eighth of the TCH bit rate (2.85 kb/s). The association of time slots and SDCCHs is rather complicated and can be found in [14].

Common Control Channels (CCCHs). One RF carrier in each cell contains a CCCH. This channel is always in time slot 0 and can be extended—if necessary—to time slots 2 and 4. The other time slots on the carrier can be used for dedicated channels.

Frames with CCCH arc organized in multiframes that are different from the TACH multiframes. They contain 51 frames (numbered 0–50) and have a duration of 235 ms. A CCCH is time-divided into a number of common (point-to-multipoint, unidirectional) channels, for signaling between a BSS and all mobiles in the cell that are active, but not involved in a call:

BCCH (Broadcast Control Channel). This is a downlink channel that broadcasts information about the network (similar to the overhead parameter messages of AMPS).

SCH (Synchronization Channel). This is a downlink channel with information that enables the mobiles to acquire frame and time-slot synchronization with the BSS.

FCCH (Frequency Correction Channel). A downlink channel with information for mobiles concerning frequency synchronization with the RF carrier.

PAGCH (Paging and Access Grant Channel). This is a downlink channel that broadcasts paging messages. Also, when the network (MSC) allocates a SDCCH to a MS, it informs the MS with a message on PAGCH.

RACH (Random Access Channel). This is the only uplink common-control channel and is used by mobiles to request a SDCCH from the network.

On the downlink of CCCH, FCCH is always in frames 0, 10, 20, 30, and 40; SCH is always in frames 1, 11, 21, 31, and 41; and BCCH is always in frames 2–5.

The sizes (number of frames) of PAGCH and RACH depend on the traffic in the cell. They are determined by the operator of the PLMN and allocated, respectively, to the remaining downlink frames and to the uplink frames.

12.7.3 Signaling Interfaces and Protocols

The interfaces and protocols for signaling between a MS and the PLMN are shown in Fig. 12.7-2 [10]. ETSI has defined two interfaces: the Um (radio) interface between a MS and the BSS, and the A (cable) interface between BSS and MSC. This is different from the U.S. cellular systems (AMPS, IS-54), where it is assumed that a MSC and its associated base stations (BSs) are supplied by the same manufacturer, and that the MSC-to-BSS interface is an internal manufacturer-specific interface.

The A-interface specifications of GSM enable the operator of a PLMN to purchase MSC and BS systems from different suppliers.

Um Interface. The signaling protocol on this interface has three layers.

Physical Layer (Layer 1). This consists of those parts of the RF channels that contain signaling channels (SACCH, FACCH, BCCH, SCH, FCCH, PAGCH, RACH, and SDCCH).

Data Link Layer (Layer 2). The protocol is known as LAPDm and is a modified version of the link access protocol for D-channels of DSS1 (Section 10.2). For details, the reader is referred to [15].

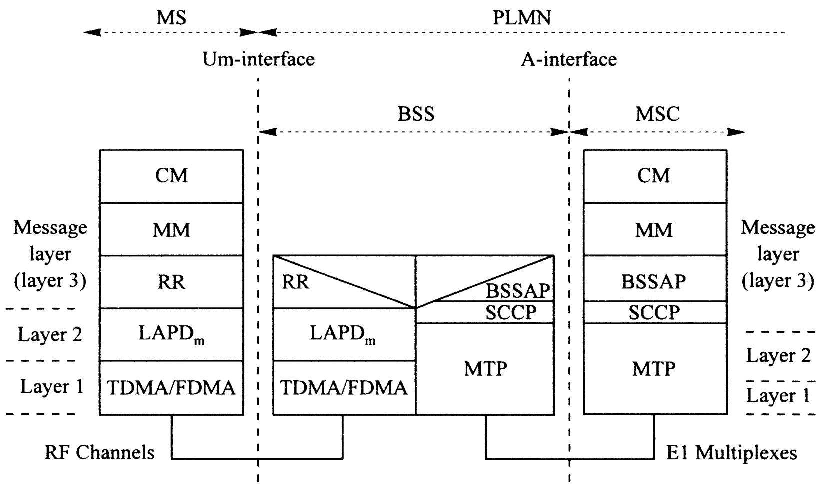


Figure 12.7-2. Um and A interfaces. (From *IEEE Commun. Mag.* **31**(4). Copyright © 1993 IEEE.)

Message Layer (Layer 3). In the MS, this layer consists of three parts. The *radio resource management* (RR) sublayer at a MS communicates with its peer in the BSS. For example, when RR at BSS allocates a TACH or a SDCCH channel to MS, it informs the MS with a RR message.

The *mobility management* (MM) and *connection management* (CM) sublayers at a MS communicate with their peers at the MSC. MM and CM messages traverse the Um and A interfaces and are transferred transparently across the BSS.

The MM sublayer messages support MS *location updating* (the GSM counterpart of "registration") and *authentication.*

The CM sublayer has three parts: *call control* (CC), *supplementary services* (SS), and *short message service* (SMS). CC contains the messages for the setup and release of connections to the MS. These messages are patterned after the Q.931 messages of DSS1 (Section 10.3). SMS is a service by which subscribers can send short (text) messages to a MS.

A Interface. The signaling protocol on this interface [16] consists of three layers that are similar to those in Signaling System No. 7.

Physical Layer (Layer 1). The BSS is connected to its MSC by digital El multiplexes (Section 1.5.2). The majority of the 64-kb/s multiplexed channels are digital trunks. During a call, a digital trunk, in tandem with a TCH channel, conveys the MS user's speech or data between the MS and the MSC [17]. Other E1 channels are signaling data links. Layers 2 and 3 of the protocol are outlined below.

Data Link Layer (Layer 2). This consists of the Message Transfer Part level 2 (MTP2) of Signaling System No. 7 (SS7)—see Chapter 8. MTP2 is responsible for the reliable transfer of signaling messages between the MSC and the BSS. Details can be found in [18].

Message Layer (Layer 3). The Base Station System Application Part (BSSAP) is present at the MSC and BSS [19]. It is a user of the signaling connection control part (SCCP) of SS7 (Chapter 15). This is one of the few applications of connection-oriented SCCP. A signaling connection is established whenever a dedicated channel (SDCCH, FACCH, or FACCH) has been assigned to a MS.

BSSAP consists of two parts: the Direct Transfer Application Part (DTAP) and the BSS Management Application Part (BSSMAP).

We start with BSSMAP at BSS. The RR at a BSS is involved in the allocation, encryption, and release of dedicated radio channels and in the transmission and reception of RR messages on the common-control radio channels. The BSSMAP at BSS and MSC handles the transfer of RR-related BSSMAP messages. At a BSS, RR and BSSMAP communicate with each other. Figure 12.7-3 illustrates a few RR–BSSMAP interactions. In example (a), a MS sends a RR message—on the RACH—to request a dedicated radio channel. RR at BSS then allocates a channel, returns a RR message on PAGCH that includes the identity of the channel, and informs its BSSMAP, which then composes a message that informs the MSC and sends it to BSSMAP at the MSC. In example (b), the MSC sends a BSSMAP message requesting the transmission of a RR message on a common-control radio channel. The requested RR message may be a paging message,

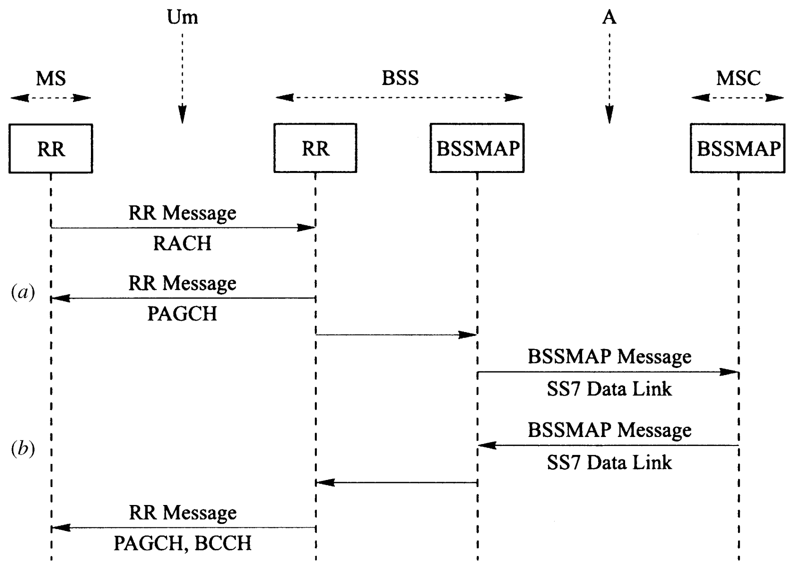


Figure 12.7-3. Transfer of RR messages. (From GSM 04.07 Version.10.0. Courtesy of ETSI.)

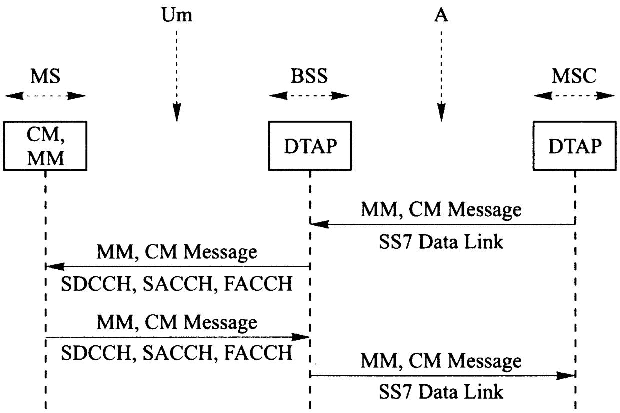


Figure 12.7-4. Transfer of MM and CM messages. (From GSM 04.07 Version 4.10.0. Courtesy of ETSI.)

which is then transmitted by BSS on the PAGCH, or a broadcast message, which is transmitted on the BCCH.

The DTAP at a BSS transparently (without processing by BSS) transfers MM and CM messages, received on dedicated radio channels, to a SS7 data link, which transports them to the MSC—see Fig. 12.7-4. Also, messages for a MS received from DTAP at the MSC are transferred to the radio channel that is currently dedicated to the MS.

The transfer of BSSMAP and DTAP messages involves the SCCP and MTP in the MS and BSS [12]. The SCCP at the sending end adds a *discrimination parameter* to the messages, which indicates whether the message belongs to BSSMAP or DTAP, and which is used by SCCP at the receiving end to deliver the message to the proper entity.

12.7.4 Identification of GSM Entities

Since GSM systems are deployed in many countries, ITU-T and ETSI have standardized the identification of GSM entities according to a numbering plan specified in ITU-T Rec. E.212 [20]. This plan is different from the ITU-T Rec. 163/164 numbering plan for fixed networks.

PLMN Identity. A PLMN is uniquely identified by its *mobile country code* (MCC) and *mobile network code* (MNC). MCC consists of three digits, of which the first one indicates a world zone:

2	Europe
3	North America
4	The Mideast and Western Asia
5	Eastern Asia and Australia
6	Africa
7	Latin America

The second and third digits represent individual nations in these zones. For example, the MCCs of the United Kingdom and Malaysia are 234 and 502.

The MNC identifies a PLMN in a country. It consists of two digits and is allocated by individual national organizations for mobile telecommunication standards.

Location Area Identity (LAI). This uniquely identifies a GSM location area. The service area of a PLMN is divided into a number of MSC service areas, and each of these is subdivided into location areas that consist of a number of adjacent cells. A MSC keeps track of the location areas of mobiles currently registered in its service area. When a MS has to be paged, paging messages are sent out in all cells of the mobile's present location area. The LAI format is

$$\text{LAI} = \text{MCC-MNC-LAC}$$

where the location area code (LAC) identifies a location area within a PLMN. The code consists of up to four hexadecimal digits. LACs are allocated by the operators of individual PLMNs.

Mobile Station Identity. We begin by pointing out a significant difference between U.S. mobiles and GSM mobiles. A GSM mobile consists of two parts: the *mobile equipment* (ME) and the *subscriber identity module* (SIM). The ME is an "almost complete" mobile station, containing the RF equipment, keypad, mouthpiece, earphone, and so on. The SIM is a small package (smart card) with semiconductor chips that store permanent and temporary information about the ME user. A user can insert her SIM into—and extract it from—any ME. The ME is operable only when a SIM has been inserted.

Splitting a MS into two parts allows a ME to be used—at different times—by different people. For example, a business may have a number of employees, each of whom has a SIM. The business also owns a smaller number of MEs which are used by the employees on a "when needed" basis.

Mobile station identities can be in two forms.

International Mobile Station Identity (IMSI). This uniquely identifies a mobile station in any GSM network. Its format is

$$\text{IMSI} = \text{MCC-MNC-MSIN}$$

where MCC-MNC identifies the PLMN selected by the MS owner for mobile services, and MSIN (mobile station identity number) identifies a MS in that PLMN. The maximum length of MSIN is nine BCD digits.

An IMSI is allocated by the operator of the selected PLMN and entered into permanent memory of the SIM.

The second MS identity format is a combination of LAI and TMSI.

Temporary Mobile Station Identity (TMSI). This is a 32-bit binary number that uniquely identifies the MS within one location area, or a group of adjacent location areas, of a PLMN. TMSI is a temporary identification and is usually changed by the network when the MS enters a new location area. LAI and TMSI are stored in temporary SIM memory.

Most messages on the Um (radio) interface identify a MS by TMSI and LAI. IMSIs are used only in exceptional cases. TMSI gives protection against cloning (obtaining MS identifications for fraudulent use), because an intercepted TMSI no longer identifies the mobile after it has left the location area.

International Mobile Equipment Identity (IMEI). This is the counterpart of the mobile serial number (MSN) in AMPS. It uniquely identifies a ME. IMEI is a 15-digit number, entered into permanent ME memory by the manufacturer.

12.8 SIGNALING BETWEEN MOBILE AND NETWORK

12.8.1 Introduction

This section explores the most important signaling procedures on the Um (radio) interface between a MS and a GSM network [21]. As has been mentioned in the previous section, some layer 3 messages on this interface are between the MS and the BSS of the cell that serves the mobile, and others are between the MS and the MSC that controls the BSS. This section describes the signaling as seen by the MS and considers the BSS and MSC as "the network."

Each message mentioned in this section points to the subsection of Section 12.9 that contains the message description. It will be helpful to read Section 12.8 twice, first ignoring the message descriptions, and looking them up the second time around.

12.8.2 MS Initialization

When a MS is powered up by its user, it needs to find the RF carrier in its cell that carries the CCCH, and then to achieve synchronization with the Broadcast Channel (BCCH), the Paging Channel (PAGCH), and the Random Access Channel (RACH), which enable the MS to listen to broadcast and paging messages, and to access the network, for location updating and originating calls.

In the AMPS/IS-54 system, a predetermined group of RF carriers has been designated by TIA as control channels, and a MS initializes by tuning to the strongest one of these channels (Section 12.1.2). GSM does not designate the RF carriers that carry common-control channels, and a GSM mobile has to scan all RF carriers operating in the cell and to find the carrier whose time slots 0 contain CCCH.

As a first step, the MS scans the downlink RF channels and looks for the Frequency Correction Channel (FCCH), whose bursts have a distinguishable pattern (148 consecutive zeros) and occur in time slots 0 of frames 0, 10, 20, 30, and 40 of the 51-frame multiframe. When the MS recognizes a FCCH burst, it knows that it is tuned to the RF carrier with CCCH, that the current time slot is slot 0, and that the current frame is one of the five listed above. The MS now initializes its (eight-step) time-slot counter and obtains time-slot synchronization.

The MS then reads the Synchronization Channel (SCH), whose bursts occur 8 BPs after the FCCH bursts. SCH information indicates which one of the above frames is the current CCCH frame. The MS then initializes its (51-step) CCCH frame counter and is now synchronized with CCCH frames. Since BCCH appears in predetermined frames (2–5) of CCCH, the MS is now also synchronized with BCCH and begins to read the System Information messages that are transmitted periodically on BCCH. There are messages that indicate the frames used for PAGCH and RACH. The MS stores this information and is now synchronized with all common-control channels in the cell.

SCH information also indicates the current frame number (0–25) in the multiframe for TACHs. This enables the MS to initialize its (26-step) frame counter, for synchronization with the TACHs.

12.8.3 Idle MS

Idle mobiles can move from cell to cell. In order to remain in contact with the network, a MS periodically compares the signal strength of the RF carrier with the CCCH in its cell to RF "CCCH" carriers in adjacent cells. The MS knows the channel numbers of these carriers from the system information messages.

If one or more of these RF carriers is stronger than the current carrier, the MS retunes to the strongest one and then reinitializes itself as described above.

12.8.4 Location Updating

GSM "location updating" is the counterpart of "registration" in AMPS and IS-54 systems. GSM networks keep track of the location area where the MS is operating [22]. When receiving an incoming call, the MS is paged in all cells of its current location area. GSM mobiles do a location update when entering a new location area, and at periodic intervals. In addition, some MSs are able to do a location update when being activated or deactivated by their users.

Updating on Entering a New Location Area. The location area identity (LAI) is broadcast in System Information messages on BCCH and stored in MS memory. When a new received LAI does not match the previously stored LAI, the MS does a location update. This happens when an idle MS has tuned to a new BCCH carrier in a different location area.

Periodic Updating. Whenever a MS does a location update, it resets a timer T. The timer has a timeout value of several hours. When the timer expires, the MS does a location update.

Updating on Deactivation and Activation. Mobiles equipped to do these updates send an IMSI DETACH message when being deactivated. The network then marks the MS as deactivated and does not send paging messages for the MS until it is activated again. A MS that has sent an IMSI DETACH message does a location update when it has been activated again.

Figure 12.8-1 shows the messages and channels for a location updating procedure. Whenever a MS has to access the network, it first requests a SDCCH, by sending a CHANNEL REQUEST message (Section 12.9.1) on RACH. The network then allocates a SDCCH and returns an IMMEDIATE ASSIGNMENT

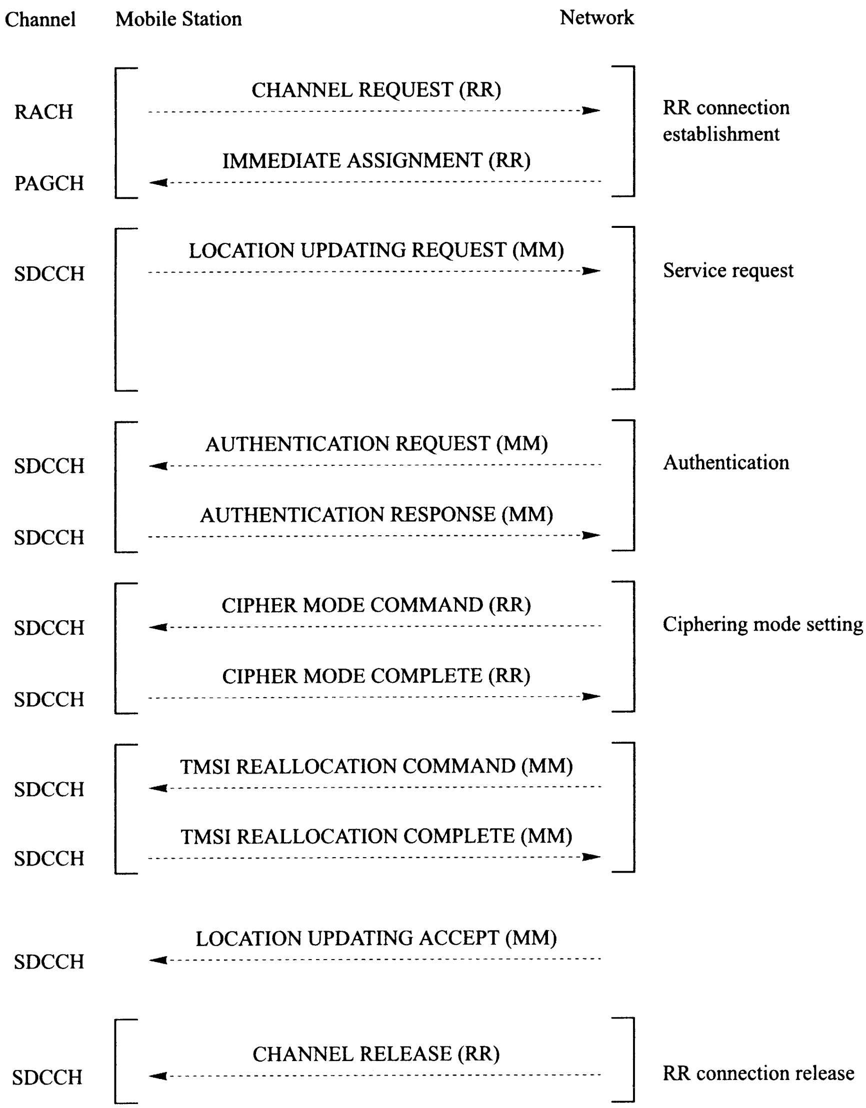


Figure 12.8-1. Location updating. (From GSM 04.08 Version 4.10.0. Courtesy of ETSI.)

message (Section 12.9.1) on PAGCH, identifying the SDCCH. From this point on, all messages are on SDCCH.

The MS sends a LOCATION UPDATING REQUEST message (Section 12.9.2), which includes information about its transmission characteristics, its identity, and indicating whether it is equipped for enciphering (data encryption).

If the SIM in the MS has been used before, its memory stores the most recent location area identity (LAI) and TMSI, and the MS identifies itself with these parameters. Otherwise, the MS uses its IMSI (stored in semipermanent SIM memory). The network checks the MS identification and—if valid—starts the MS authentication.

The GSM authentication algorithm has two inputs: a random number RAND (128 bits) and an *individual subscriber authentication key* Ki (128 bits). The key is stored in the SIM of the mobile and in the network. It is never transmitted on the Um (radio) interface. The MS and the network execute the algorithm, using the same RAND and their respective Ki values. The results of the algorithm are the *signed result* SRES (32 bits) and the *cipher key)* Kc (64 bits).

The network starts the authentication by sending an AUTHENTICATION REQUEST (Section 12.9.2), which includes a RAND value. The mobile executes algorithm and returns an AUTHENTICATION RESPONSE (Section 12.9.2) that includes SRES. The network compares the received SRES with the SRES it has calculated. If they match, the network considers the MS as authenticated and continues the location updating procedure.

If the MS has indicated that it can handle ciphered (encrypted) information, the network then sends a CIPHER MODE COMMAND message (Section 12.9.1), and the MS responds with a CIPHER MODE COMPLETE (Section 12.9.1) message. From this point on, the MS and BSS encrypt their information on SDCCH, using the value of Kc to form an encryption/decryption mask. If the MS has indicated that it cannot handle ciphering, this step is omitted.

The network may decide to allocate a new TMSI to the mobile. In this case, it sends a TMSI REALLOCATION COMMAND (Section 12.9.2) that includes the new TMSI. After entering the TMSI in its SIM, the MS sends a TMSI REALLOCATION COMPLETE message (Section 12.9.2), and the network sends a CHANNEL RELEASE (Section 12.9.1). It then releases SDCCH, and the MS also releases the channel.

If neither the authentication nor TMSI reallocation is done, the network sends a LOCATION UPDATING ACCEPT message (Section 12.9.2) to inform the MS and then sends the CHANNEL RELEASE.

12.8.5 Setup of a Call Originated by MS

When the user of a MS originates a call, he first enters the called number and.possibly additional information with the MS keypad, and then depresses the “send” button Figure 12.8-2 show the signaling for the setup of the call.

The initial signaling is very Similar to the initial signaling for location updating. The MS first requests a SDCCH with a CHANNEL REQUEST message

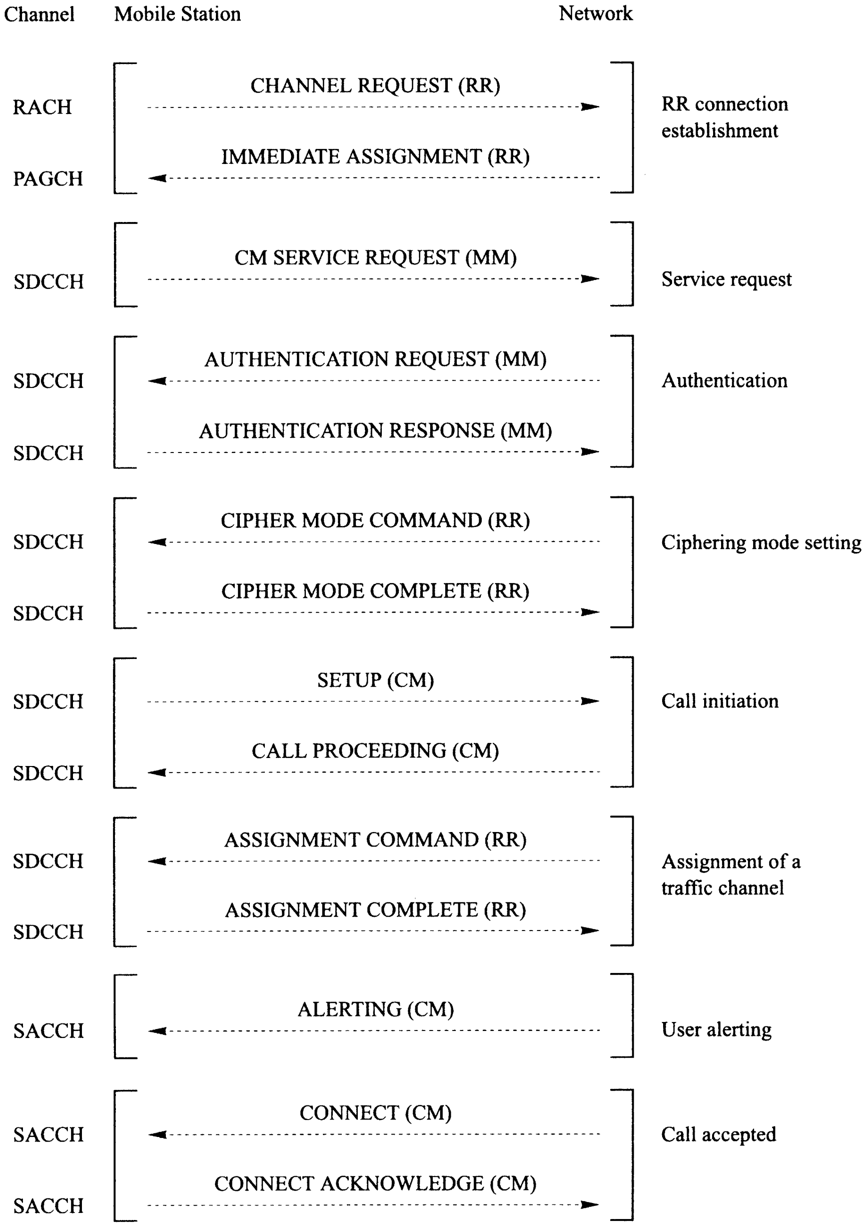


Figure 12.8-2. Call originated by MS. (From GSM 04.08 Version 4.10.0. Courtesy of ETSI.)

(Section 12.9.1). The network then assigns a SDCCH and returns an IMMEDIATE ASSIGNMENT message (Section 12.9.1). At this point, the MS and BSS tune their transmitter and receiver to the SDCCH.

The MS next sends a CM SERVICE REQUEST (Section12.9.2), indicating that it wishes to originate a call. The network then decides whether to execute or skip the MS authentication. If the MS has ciphering capability, they ciphering mode is started by command from the network.

The call-control messages that follow are similar to the corresponding Q.931messages of Digital Subscriber Signaling System No. 1 (Section 10.3.2).

The MS sends a SETUP message (Section 12.9.3) that includes the called party number, bearer capability, and other information. The network acknowledges with a CALL PROCEEDING message (Section 12.9.3). It then allocates a TACH, sends an ASSIGNMENT COMMAND (Section 12.9.1), and then releases SDCCH. On receipt of the command, the MS sends an ASSIGNMENT COMPLETE (Section 12.9.1) message tunes to TACH. From this point, the signaling is on the SACCH to TACH.

When the network receives an indication from the fixed network that the called party is being alerted, it sends an ALERTING message (Section 12.9.3), and the MS notifies its user.

When the called party answers, the network sends a CONNECT message (Section 12.9.3), and the MS returns a CONNECT ACKNOWLEDGE (Section 12.9.3). The conversation can now begin.

12.8.6 Connection Setup for Call Terminating at MS

In Fig. 12.8-3, a MS is active and idle, and listening to the P AGCH of its cell. When the mobile network receives a call for the MS, it transmits PAGING REQUEST in all cells of the MS location area (Section 12.9.1). The called MS is identified by TMSI or IMSI. When recognizing that it is being paged, the MS sends a CHANNEL REQUEST (Section 12.9.1) message. The network then allocates a SDCCH and returns an IMMEDIATE ASSIGNMENT message (Section 12.9.1).

The MS acknowledges the paging request with a PAGING RESPONSE message (Section 12.9.1). The network then decides Whether to execute or skip the MS authentication. If the MS has ciphering capability, the network initiates the ciphered mode on SDCCH.

The message sequence that follows is similar to that of Section 12.8.5, except that for some messages the roles of the MS and the network are reversed. The network starts with a SETUP message (Section 12.9.3), and the MS acknowledges with a CALL CONFIRMED message (Section 12.9.3).

At this point, the network allocates a TACH and sends an ASSIGNMENT COMMAND (Section 12.9.1). The MS acknowledges with an ASSIGNMENT COMPLETE (Section 12.9.3.1) message and then tunes to TACH. The network then releases SDCCH, and signaling for the rest of the call is on the SACCH or FACCH of TACH.

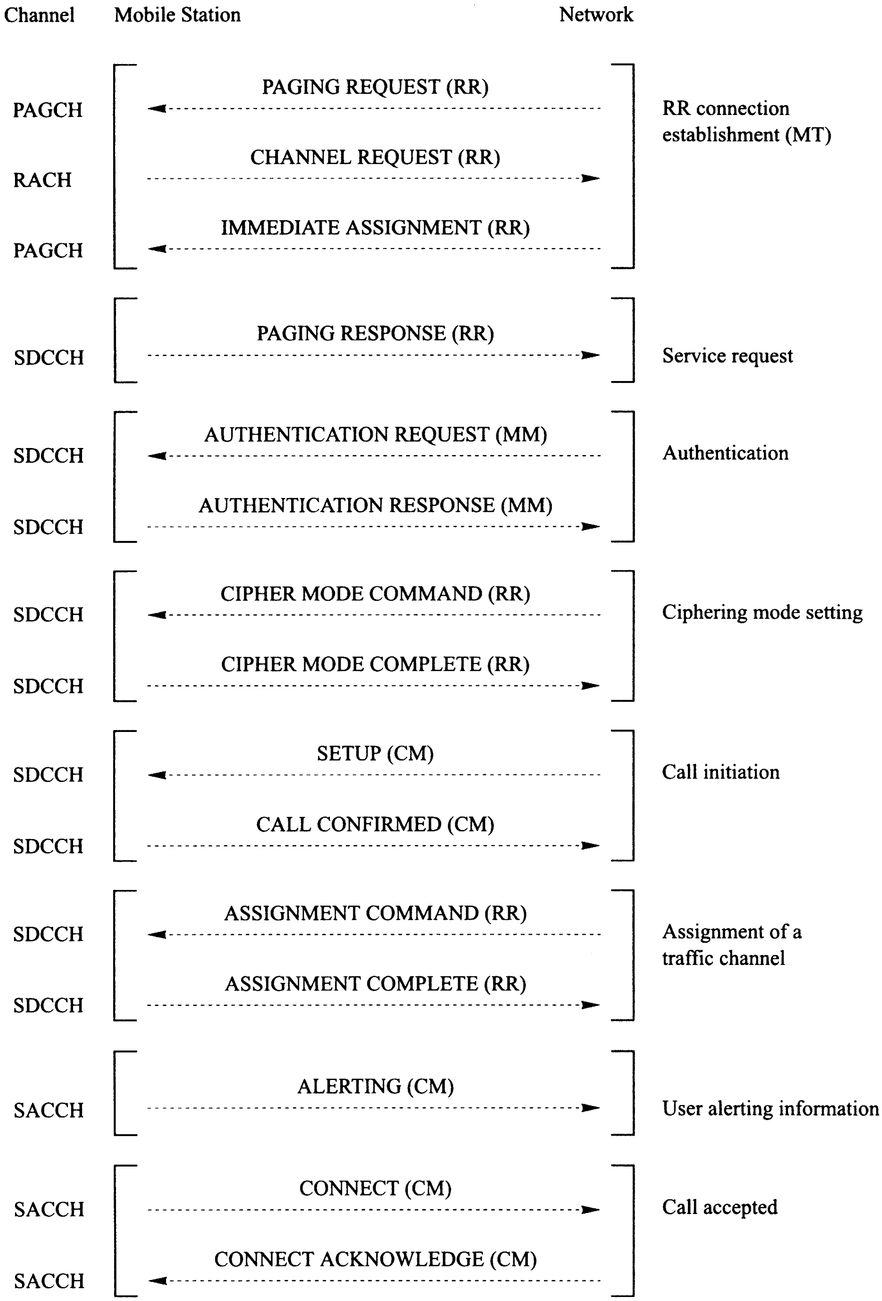


Figure 12.8-3. Call terminating at MS. (From GSM 04.08 Version 4.10.0. Courtesy of ETSI.)

The MS sends an ALERTING message (Section 12.9.3) when it starts to alert its user. When the user answers, the MS sends a CONNECT message (Section 12.9.3), the network returns a CONNECT ACKNOWLEDGE (Section 12.9.3), and the conversation starts.

12.8.7 Release of Connection

The release of a connection can be initiated by a disconnect message from the MS or (via the network) by a disconnect message or signal from the other party. Figure 12.8-4 shows the message sequences for both cases. The DISCONNECT message (Section 12.9.3) indicates that the message originator has disconnected. The RELEASE message (Section 12.9.3) acknowledges the receipt of the DISCONNECT and indicates that its originator will stop sending and receiving on TACH.

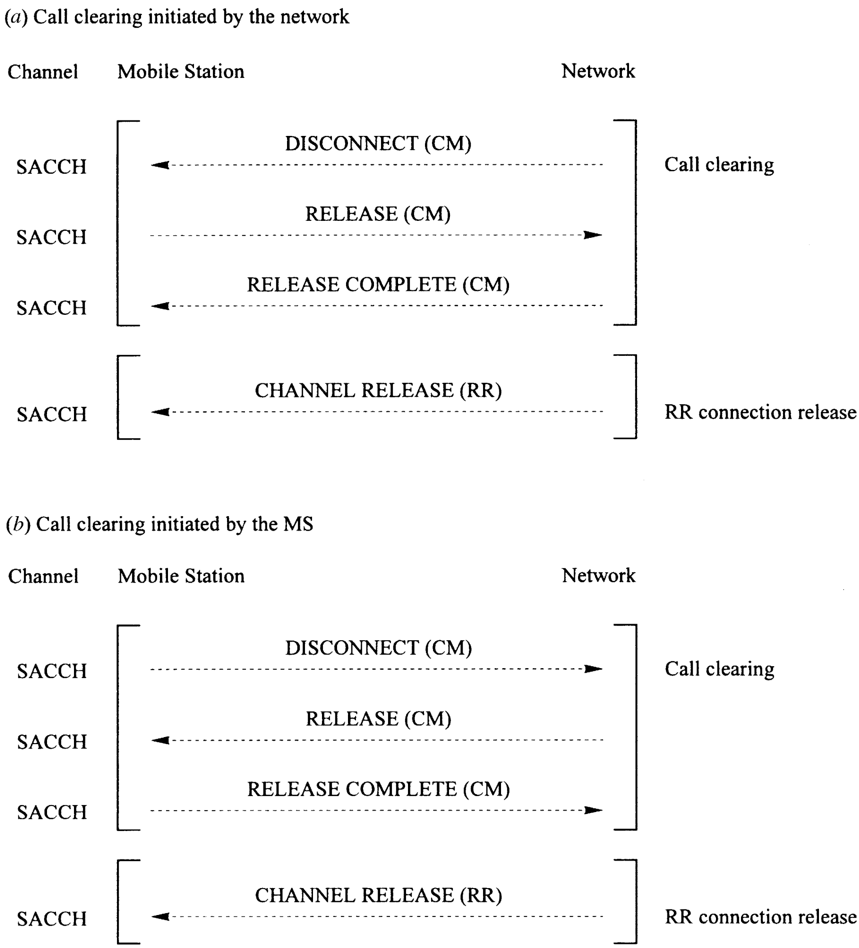


Figure 12.8-4. Call clearing. (From GSM 04.08 Version 4.10.0. Courtesy of ETSI.)

The RELEASE COMPLETE message (Section 12.9.3) indicates that the other party also will stop sending and receiving information.

After the network has sent or received the RELEASE COMPLETE, it sends a CHANNEL RELEASE message (Section 12.8.1) and then releases TACH. On receipt of the message, the MS also releases TACH.

12.9 LAYER 3 MESSAGES ON THE Um INTERFACE

ETSI has defined some seventy GSM layer 3 messages on the Um interface. They are grouped according to the sublayers at the MS (Fig. 12.7-2). This section describes the messages mentioned in the previous section.

Figure 12.9-1 shows the general message layout. The *protocol discriminator* specifies the message as a RR, MM, or CM message. Within each protocol, the particular message is indicated by the *message type*. Messages can include mandatory and optional *information elements* IEs. The descriptions include the most important IEs in the messages. For other IEs and details of IE coding, see [21].

12.9.1 Messages for Radio Resources Management (RR)

Assignment Command. This message is sent (on SDCCH) from the network to the MS and indicates that a TACH has been assigned. Included is the *channel description* IE, which specifies the RF carrier channel number (ARFCN) and time slot of TACH. Also included is the *channel mode* IE, which indicates the communication mode on the channel (signaling, speech, or data).

Assignment Complete. This is a response to the assignment command, sent by the MS (on SDCCH). It indicates that the mobile will tune to the assigned channel.

Channel Release. This message is sent by the network—on the SDCCH or TACH that is currently assigned to the MS. It indicates that the network has released the channel. It includes the *release cause* IE, which gives the reason for the release (normal release, release caused by failure, etc.).

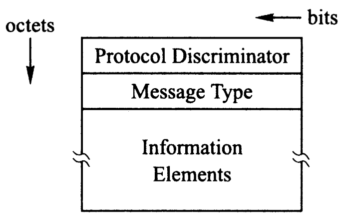


Figure 12.9-1. Layout of layer 3 messages on the Um interface. (From GSM 04.08 Version 4.10 0. Courtesy of ETSI.)

Channel Request. This message is sent (on RACH) by a MS, which currently has no dedicated channel (SDCCH to TACH), or request a SDCCH. For security, the MS does not directly identify itself, but generates a random *reference number* IE and includes it in the message.

Cipher Mode Command. This message is sent by the network (on SDCCH). It indicates that the network will encipher and decipher (encrypt and decrypt) all subsequent information on SDCCH and TACH and requests the MS to do the same.

Cipher Mode Complete. This message is sent by the MS (on SDCCH) to acknowledge the above command.

Immediate Assignment. This message is sent by the network (on PAGCH) in response to a channel request message. It indicates that a SDCCH has been allocated and requests the MS to tune to this channel. The message includes the *channel description* IE, which indicates the absolute RF carrier number (ARFCN), the time-slot number (0–7) and, since a SDCCH is not full-rate channel, information regarding the frames in which the channel appears. Also included is the *request reference* IE, which is equal to reference number received in the channel request message. A MS accepts an assignment message when it recognizes the reference number of its channel request message.

Paging Request. This message is sent by the network (on PAGCH). It informs a MS about an incoming call and includes the TMSI of the called MS.

Paging Response. This message is sent by the MS (on SDCCH) to confirm the receipt of a paging request. It includes TMSI and the *MS class mark* IE. The latter IE specifies the mobile's RF transmitter power class and indicates whether the mobile has ciphering capability.

System Information Messages. These messages are broadcast by the network on BCCH. There are several information messages. Examples of included IEs are:

- Location area identity (LAI). This identifies the location area of the cell that is broadcasting the message.
- Cell options. This indicates whether the MS is allowed to use discontinuous transmission (DTX, see Section 12.4.2).
- RACH control. This indicates the maximum number of allowed MS attempts to access RACH and the number of burst periods (BPs) between successive attempts.
- Neighbor cells description. This lists the ARFCNs of RF channels in neighboring cells that carry broadcast channels (BCCHs).

12.9.2 Messages for Mobility Management (MM)

Authentication Request. This message is sent by the network (on SDCCH) to check the authenticity of the MS. It includes the random number IE (RAND).

Authentication Response. This message is sent by a MS (on SDCCH) in response to the Authentication Request. It includes the *signed response* (SRES).

CM Service Request. This message is sent by a MS (on SDCCH) to request a connection management (CM) service. It includes the *CM service type* IE, which specifies the requested service (e.g., the setup of a circuit-switched connection, the activation of a supplementary service). Also included are the MS class mark and TMSI.

IMSI Detach Indication. This message is sent by a MS (on SDCCH) just before it powers down. It indicates that the MS cannot receive new incoming calls. It includes the MS class mark and TMSI.

Location Updating Accept. This message is sent by the network (on SDCCH) to indicate acceptance of the Location Updating Request. It includes the location area of the cell that sends the message.

Location Updating Request. This message is sent by a MS (on SDCCH). It includes the MS class mark and the IMSI, or TMSI, and the location area (LAI) stored in the SIM of the mobile.

TMSI Reallocation Command. This message is sent by the network to inform the MS that a new TMSI has been allocated. It includes TMSI and the location area of the cell that sends the message.

TMSI Reallocation Complete. This message is sent by a MS, acknowledging that it has received the reallocation command and has entered the data into its SIM.

12.9.3 Messages for Circuit-Switched Call Control (CC)

We only consider messages in the CC part of the connection management (CM) layer. GSM intends to provide both PSTN and ISDN services, and the CC messages are the counterparts of the Q.931 messages of digital subscriber signaling system No. 1 (DSS1—see Section 10.3.2). The information elements are also patterned after those of DSS1 (Section 10.3.4) but have been adapted for mobile communications. All messages are sent on a control channel of the TACH that is currently dedicated to the MS.

Alerting. On calls originated by a MS, the message is sent from the network and indicates that the called party is being alerted. On calls to a MS, the message is sent by the MS and indicates that it is alerting its user.

Call Confirmed. This message is sent by a called MS, indicating that it is processing the received Setup message. The message has no counterpart in Q.931.

Call Proceeding. This message is sent by the network, on calls originated by a MS. It acknowledges the Setup message and indicates that the connection is being set up.

Connect. On calls originated by a MS, the message is sent by the network. On calls terminating at a MS, it is sent by the MS. It indicates that the called party has answered.

Connect Acknowledge. This message acknowledges the receipt of a connect message.

Disconnect. This indicates that the MS user, or the other party in the call, has disconnected. It requests the release of the connection. The message is sent by the MS or the network, depending on which party has cleared. It includes the *cause* IE, which indicates the disconnect reason (e.g., normal call clearing).

Release. This message is sent in two situations. It either acknowledges a received disconnect message or indicates a network-initiated release. In the latter case, the message includes a *cause* IE, which indicates why the call setup has failed: for example, unassigned called number received, called party busy, or network failure. The message indicates that the originating entity is about to release the connection at its end.

Release Complete. The message acknowledges a Release message. It indicates that its originator will release the connection at its end.

Setup. This message is sent by a MS or the network on, respectively, calls originating or terminating at the MS. The IEs that may be included are outlined below.

- Bearer capability. Mandatory in direction MS → MSC; optional in the other direction. This is a compound IE, which includes several information items, for example:

  Radio channel requirement (full rate, half rate).

  Information transfer capability (speech, unrestricted digital information, 3.1-kHz audio, group 3 facsimile).
- Called party address. Mandatory in direction MS → MSC; optional in the other direction.
- Calling party address. Optional in both directions.
- Called and calling party subaddresses. Optional in both directions.
- High- and low-layer compatibility. Optional in both directions.
- User–user information. Optional in both directions.

12.10 ACRONYMS

ACT	Action field
AMPS	Advanced mobile phone service
AUC	Authentication center
AUTH	Authentication indicator
AUTHR	Result of authentication calculation
BCCH	Broadcast Control Channel
BCD	Binary coded decimal
BI	Busy/idle bit
BIS	Busy/idle status indicator
BP	Burst period
BS	Base station
BSS	Base station subsystem
BSSAP	Base Station System Application Part
BSSMAP	Base Station System Management Application Part
BST	Base station trunk
BSTG	Base station trunk group
CAVE	Cellular authentication and voice encryption
CC	Call control
CCCH	Common-Control Channel
CDMA	Code-division multiple access
CEPT	European Conference of Postal and Telecommunications Administrations
CFW	Control filler word
CHAN	Channel number
CM	Connection management
CMAC	Mobile attenuation code on control channel
CMAX	Number of access channels
CMN	Cellular mobile network
CPA	Combined paging and access indicator
DCC	Digital color code
DL	Data link
DOT	Dotting (synchronization) sequence
DQPSK	Differential quaternary phase-shift keying
DSS1	Digital subscriber signaling system No. 1
DTAP	Direct transfer application part
DTX	Discontinuous transmission
DVCC	Digital verification color code
E	Extended address indicator
EIA	Electronic Industries Alliance
ETSI	European Telecommunications Standards Institute
FACCH	Fast Associated Control Channel
FCCH	Frequency Correction Channel
FOCC	Forward Control Channel

FVC	Forward Voice Channel
GAW	Global action word
GSM	Global System for Mobile Communications
IMEI	International mobile equipment identity
IMSI	International mobile station identity
IS-54	U.S. standard for TDMA mobile system
IS-95	U.S. standard for CDMA mobile system
ISDN	Integrated Services Digital Network
ITU	International Telecommunication Union
ITU-T	Telecommunication Standardization Sector of ITU
Kc	Cipher key
Ki	Individual subscriber authentication key
LAC	Location area code
LAI	Location area identity
LT	Last try indicator
MAC	Mobile attenuation code
MCC	Mobile country code
MIN	Mobile identification number
MIN1	Exchange code and line number of MIN
MIN2	Area code of MIN
MM	Mobility management
MNC	Mobile network code
MPCI	Mobile protocol capability indicator
MS	Mobile station
MSC	Mobile switching center
MSCTG	Trunk group between MSCs
MSG TYP	Message type
MSIN	Mobile subscriber identity number
MSN	Mobile serial number
N	Number of paging channels
NAWC	Number of additional words coming
OHD	Overhead field
OPM	Overhead Parameter message
ORD	Order field in mobile control messages
ORDQ	Order qualifier field in mobile control messages
PACH	Paging and Access Grant Channel
PIN	Personal identification number
PLMN	Public land mobile network
PSCC	Color code of present voice channel
PSTN	Public switched telecommunication network
PW1	First word of overhead parameter message
PW2	Second word of overhead parameter message
RACH	Random Access Channel
RAND	Random number

RANDC	Eight leading bits of RAND
RCF	Read control filler indicator
RECC	Reverse Control Channel
REGH	Registration indicator for home mobiles
REGID	Registration ID
REGINCR	Registration increment
REGR	Registration indicator for roaming mobiles
RF	Radiofrequency
RR	Radio resource management
RSVD	Reserved message field
RVC	Reverse voice channel
SACCH	Slow Associated Control Channel
S	MSN indicator
SACCH	Slow Associated Control Channel
SAT	Supervisory audio tone
SCH	Synchronization Channel
SCM	Station class mark
SDCCH	Stand-Alone Dedicated Signaling Channel
SID	System identification (15 bits)
SID1	14 leading bits of SID
SRES	Signed result
SSD_A	Shared secret data for authentication
SSD_B	Shared secret data for voice privacy
ST	Signaling tone
TACH	Traffic and Associated Control Channel
TCH	Traffic Channel
TDMA	Time-division multiple access
TG	Trunk group
TIA	Telecommunications Industry Association
TMSI	Temporary mobile station identity
VC	Voice channel
VMAC	Voice channel mobile attenuation code
WFOM	Indicator: wait for overhead message
WS	Word synchronization pattern

12.11 REFERENCES

1. Z. C. Fluhr and P. T. Porter, Advanced Mobile Phone Service: Control Architecture, *Bell Syst. Tech. J.*, **58**, No. 1, Jan. 1979.
2. K. J. S. Chadha, C. F. Hunnicutt, S. R. Peck, and J. Tebes, Jr., Advanced Mobile Phone Service: Mobile Telephone Switching Office, *Bell Syst. Tech. J.*, **58**, No. 1, Jan. 1979.
3. W. C. Y. Lee, *Mobile Cellular Telecommunication Systems*, 2nd ed., McGraw-Hill, New York, 1995.
4. *Public Land Mobile Network: General Aspects*, Rec. Q.1001, ITU-T, Geneva, 1989.

5. *Public Land Mobile Network: Network Functions*, Rec. Q.1002, ITU-T, Geneva, 1989.
6. *Mobile Station—Land Station Compatibility Specification*, EIA/TIA-553, Electronic Industries Association, Washington, DC, 1989.
7. *Cellular Features Description*, TIA/EIA/IS-53, Telecommunications Industry Association, Washington, DC, 1991.
8. D. J. Goodman, Trends in Cellular and Cordless Communications, *IEEE Commun. Mag.*, **29**, No. 6, June 1991.
9. J. E. Padgett *et al.*, Overview of Wireless Communications, *IEEE Commun. Mag.*, **33**, No. 1, Jan. 1995.
10. M. Rahnema, Overview of the GSM System and Protocol Architecture, *IEEE Commun. Mag.*, **31**, No. 4, Apr. 1993.
11. D. D. Falconer, F. Adachi, and B. Gudmundson, Time Division Multiple Access Methods for Wireless Personal Telecommunications, *IEEE Commun. Mag.*, **33**, No. 1, Jan. 1995.
12. *Cellular System Dual-Mode Mobile Station–Base Station Compatibility Standard*, EIA/TIA/IS-54-B, Telecommunications Industry Association, Washington, DC, 1992.
13. *Mobile Station–Base Station Compatibility Standard for Dual-Mode Wideband Spread Spectrum Cellular System*, TIA/EIA/IS-95, Telecommunications Industry Association, Washington, DC, 1993.
14. M. Mouly and M. B. Pautet, *The GSM System for Mobile Communications*, published by the authors, France, 1992.
15. *European Digital Cellular Telecommunication System (Phase 2), MS–BSS Interface Data Link Specification*, **GSM 04.06**, European Telecommunications Standards Institute, Sophia Antipolis, France, 1994.
16. *European Digital Cellular Telecommunication System (Phase 2), BSS–MSC Interface Principles*, **GSM 08.02**, European Telecommunications Standards Institute, Sophia Antipolis, France, 1994.
17. *European Digital Cellular Telecommunication System (Phase 2), BSS–MSC Interface Layer 1 Specification*, **GSM 08.04**, European Telecommunications Standards Institute, Sophia Antipolis, France, 1994.
18. *European Digital Cellular Telecommunication System (Phase 2), Signaling Transport Mechanisms Between BSS and MSC*, **GSM 08.06**, European Telecommunications Standards Institute, Sophia Antipolis, France, 1994.
19. *European Digital Cellular Telecommunication System (Phase 2), MSC to BSS Layer 3 Specification*, **GSM 08.08**, European Telecommunications Standards Institute, Sophia Antipolis, France, 1994.
20. *Identification Plan for Land Mobile Stations*, Rec. E.212, ITU-T, Geneva, 1989.
21. *European Digital Cellular Telecommunication System (Phase 2), Mobile Radio Interface Layer 3 Specification*, **GSM 04.08**, European Telecommunications Standards Institute, Sophia Antipolis, France, 1994.
22. *Public Land Mobile Network: Location Registration Procedures*, Rec. Q.1003, ITU-T, Geneva, 1989.

13

AIR INTERFACE SIGNALING IN CDMA NETWORKS

13.1 INTRODUCTION

CDMA, or *code-division multiple access*, is a digital wireless technology used originally for military applications because of its superior anti-jamming and anti-eavesdropping characteristics. It was developed for the civilian market by Qualcomm in the 1980s and commercialized in 1995. CDMA derives its name from the fact that all communication channels use the same RF carrier band and are separated from each other by coding. Although initially competing with TDMA, CDMA's superior bandwidth efficiency has made it the technology of choice for third-generation wireless systems worldwide.

13.1.1 Evolution of CDMA Technology

This chapter discusses three public wireless network standards for the *air interface* between base stations (BSs) and mobile stations (MSs): IS-95 and two standards collectively known as IMT-2000. Network interfaces are discussed in Chapters 12 and 19.

IS-95 CDMA. IS-95, also known by the marketing name of cdmaOne™, is the standard that was used for the first commercial CDMA systems. It is named after the Interim Standard 95 document, issued jointly by the Telecommunications Industry Association (TIA) and the Electronic Industry Alliance (EIA). IS-95 systems are part of the second-generation (2G) cellular systems introduced in the 1990s, which support voice and narrowband data [1–3].

IS-95 has two versions (*revisions*): IS-95A, the initial version, and IS-95B, which superseded it. This chapter is based on IS-95B.

Signaling in Telecommunication Networks, Second Edition, by John G. van Bosse and Fabrizio U. Devetak

IMT-2000 (International Mobile Telecommunications-2000). IMT-2000 is the ITU-T-defined umbrella standard for third-generation (3G) cellular systems that support voice and wideband data. Under IMT-2000, two sets of standards cover the radio access portion (the air interface): cdma2000® and UTRAN.

cdma2000®. This is the 3G evolution of IS-95, with which it can be made backward compatible. Wideband functionality can be deployed in steps; the main ones are called 1X and 3X (Section 13.3.1). 1X systems, sometimes called second-generation-and-a-half (2.5G) systems, use the same bandwidth as IS-95 and offer a somewhat limited improvement in data rates. 3X systems use three times the bandwidth of IS-95 and offer true wideband data performance.

UTRAN (*UMTS Terrestrial Radio Access Network*). Also called UTRA or WCDMA, UTRAN is the air interface for the Universal Mobile Telecommunication System (UMTS). UMTS is the overall standard that covers both network and air interfaces and is the 3G evolution of GSM (Section 12.7). It shares network architecture and network interfaces with GSM but uses CDMA technology for radio access. Since the radio interface is based on CDMA, UTRAN cannot be backward compatible with GSM. To address possible RF-band constraints and to support phased introductions, UTRAN allows two bandwidth configurations: *frequency-division duplex* (FDD) and *time-division duplex* (TDD), discussed in Section 13.4.1.

IS-95, cdma2000, and UTRAN systems are covered in Sections 13.2, 13.3, and 13.4, respectively.

Before describing CDMA signaling, we need to introduce some important CDMA concepts: coding is discussed in the next section, and other aspects are discussed in the section after the next. An additional concept, the *OSI layered model*, which is the framework for CDMA signaling protocols, is discussed in Section 20.1.

13.1.2 CDMA Coding

Spread Spectrum. CDMA belongs to a class of transmission methods called *spread spectrum* after the fact that the *baseband* data from the source is spread over a much wider bandwidth than the minimum needed for communication. The use of a wider bandwidth is compensated by the fact that multiple users can access that bandwidth concurrently, without interfering with each other. Different methods exist within the scope of spread spectrum, such as *frequency hopping* (FH) and *direct sequence* (DS). The CDMA systems described in this chapter use the direct sequence method [4–6].

Direct Sequence. With this method, the source's baseband digital signal is coded by cyclically repeating binary codes, called *chipping codes* because the bits in the codes are called *chips.* Code sequences are generated continuously in both mobile stations and base stations and run at a much higher rate than the baseband signal: a typical chipping code rate is 1.2288 Mc/s (megachips per second) while typical baseband user data rates are in the 9–14 kb/s range. The coding operation thus *spreads* the source signal across the wider bandwidth of the chipping code. The ratio between the chip rate and the baseband bit rate is called *processing gain*, and typical values are above 80.

The key property of CDMA chipping codes is that a user signal, coded with a particular code and transmitted by the sender, can be recovered at the receiving end by decoding (*despreading*) it with the same code, even when sharing the RF band with transmissions meant for other receivers. User signals spread by a code other than the one used for despreading produce either none or a small amount of interference. For example, assume that a BS is involved in three simultaneous connections and transmits user signals to MS1, MS2, and MS3, spread by chipping codes C1, C2, and C3, respectively. Each MS receives the sum of the three user signals, because it is tuned to the same RF band, but is able to extract its intended signal by decoding the composite signal with its particular chipping code.

Coding and decoding operations are conceptually *multiplications* (performed at the rate of the chipping codes) with binary 1's coded as −1's and binary 0's coded and +1's. In some stages of the process, multiplication may be implemented as *modulo 2 addition* (exclusive OR) on standard binary data (1's and 0's), which yields equivalent results. The next section describes coding and decoding as multiplication.

Coding/Decoding Example. Figure 13.1-1 shows coding for transmissions from sources (a) and (b), and decoding at the destination for (a). Sa (10100) and Sb (11001) are the source signals; Ca (0101) and Cb (0011) are the corresponding chipping codes. SPa and SPb are the spread signals, and TS is the sum of SPa and SPb (composite signal) received at the destination for (a). At the destination, TS is decoded with Ca, producing signal Da. Da is integrated (added cumulatively chip-by-chip) over each chipping code cycle, producing signal ISa. ISa is then converted to binary values, resulting in the received signal bISa, which is equal to Sa as intended.

Several remarks about this example are in order:

1. Signals SPa and SPb are assumed to be the same at the transmitter(s) and at the receiver; that is, the transmitted signals are received accurately.
2. Signals SPa and SPb are assumed to arrive at the destination for (a) with equal strength. The importance of this assumption and the manner in which it is realized in CDMA systems are discussed in the next section.
3. Chipping codes are assumed to be synchronized at the sources and the destinations. The reader is invited to explore the case where Ca is applied out of sync by even one chip between source and destination.
4. For simplicity, the example uses chipping codes with a period that is 1/4 the bit period of the source, thus "spreading" the source's (baseband) signal by a factor of four.
5. The codes used have properties that prevent them from interfering with each other. The general requirements for chipping codes are discussed below.

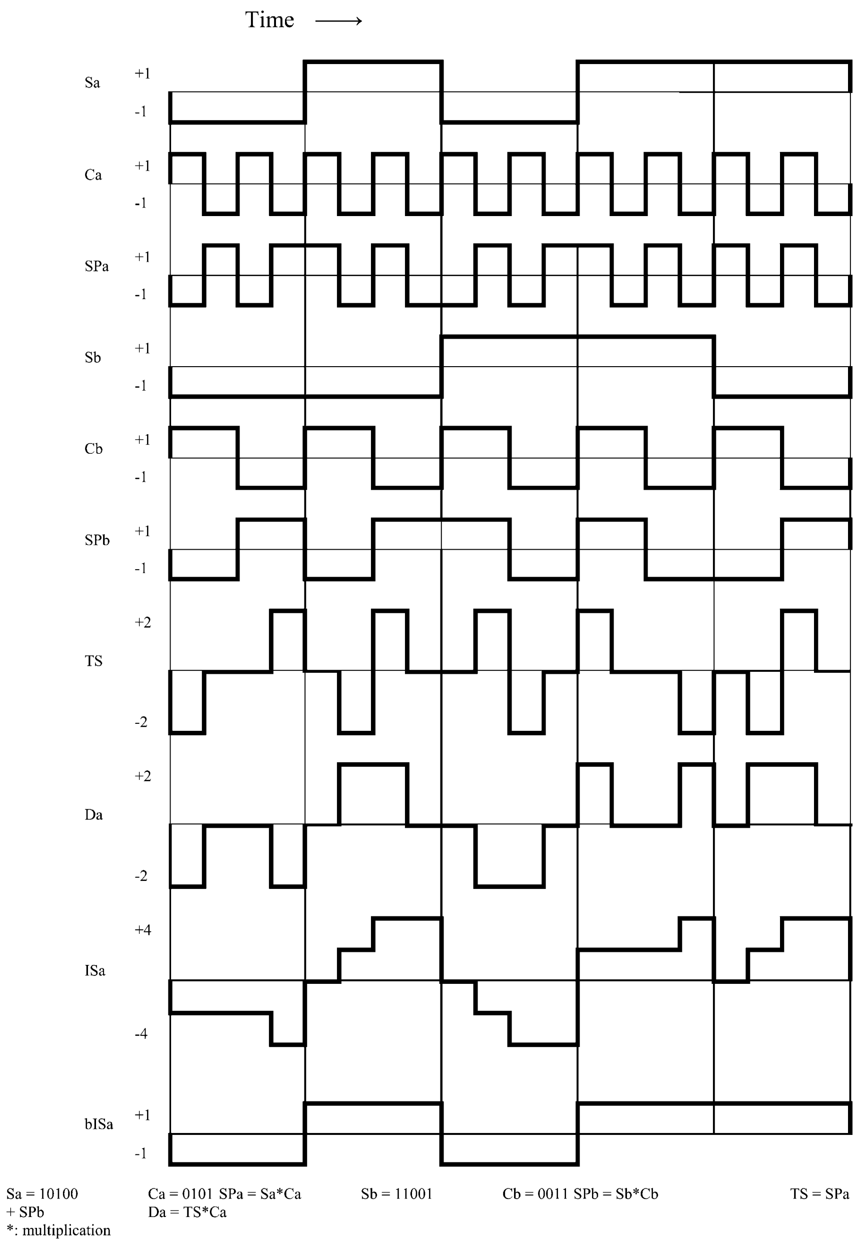


Figure 13.1-1. Example of CDMA coding and decoding.

Chipping Codes. These codes must have specific characteristics in order to separate concurrent transmissions in the same RF band. The first characteristic is that the number of 1's and 0's in a code should be the same or differ at most by one. The other characteristics are related to functions called *cross-correlation* and *autocorrelation.*

Correlation is the numerical value that results from multiplying bit-by-bit two binary sequences of equal length (encoded with bipolar values $+1$ and -1) then adding the bits of the product. The same result can be obtained with standard binary values (1's and 0's) by comparing the two sequences bit-by-bit and then subtracting the number of mismatches from the number of matches. In mathematical terms, correlation is the *scalar* that results from the *dot product* (or *inner product*) of the *vectors* that represent the two sequences.

Cross-correlation is the correlation between two different sequences; autocorrelation is the correlation between two instances of the same sequence that are phase-shifted (in a circular fashion) by a whole number of chips. For instance, if we compare the two code sequences of Fig. 13.1-1 (0101 and 0011), we find two matches (the first and the last bit) and two mismatches (the two middle bits), so the cross-correlation is $2 - 2 = 0$. If we then take the first sequence (0101) and correlate it with an instance of itself that has been shifted to the right by one chip (1010), we have zero matches and four mismatches, yielding an autocorrelation of $0 - 4 = -4$. Autocorrelation when in-phase is $4 - 0 = 4$.

A set of equal-length codes that have cross-correlation values of zero (called *orthogonal codes* because the corresponding vectors are orthogonal) can be used to separate communication flows in the same RF band, as long as sender and receiver are in-phase. *Walsh codes* (see below) are an example of orthogonal codes.

With other types of code (PN codes, see below) autocorrelation properties can be used to separate communication flows, by spreading and despreading source signals with different phase-shifted instances of the same code sequence. The key property of these codes is that the autocorrelation value when two instances of the sequence are in-phase is much higher than any of the out-of-phase values. These code types also provide a means of acquiring code synchronization between source and receiver: the source transmits a known pattern spread with a known chipping code; the receiver then despreads the received transmission with different instances of the same chipping code, each shifted progressively by a whole number of chips, until the known pattern is detected. That indicates that the sender and receiver are in-phase.

Code Types. CDMA systems use two types of codes: *Walsh codes* (also called *Hadamard codes*) and *pseudorandom noise* (PN) *codes*.

Walsh codes. These codes are relatively short (from 2 to 512 chips) and come in orthogonal sets generated from a square matrix (*Hadamard matrix* [1,4,5], so the number of codes in a set is equal to their length. They provide the best performance in terms of number of channels that can be supported in a cell, but maintain their orthogonality only when source and receiver are phase-synchronized. If sender and receiver are out of phase by even one chip, orthogonality may be lost. To ensure synchronization, the sender must transmit a *pilot channel*, which is used by the receiver to acquire the sender's phase but causes higher power consumption by the sender's transmitter. The codes in Fig. 13.1-1 are from the set of 4-bit Walsh codes: 0-0-0-0, 0-1-0-1, 0-0-1-1, 0-1-1-0.

Walsh codes are typically used to create individual channels (*code channels*) within broader channels created by PN codes.

PN codes. These codes are much longer than Walsh codes and, although generated deterministically, they display characteristics of randomness. They are not perfectly orthogonal, but their cross-correlation is close to zero and they have excellent autocorrelation properties. The phase shift between instances of the same PN sequence, called *phase offset* and expressed in number of chips, can be used to separate channels using *correlator circuits*, even when source and destination are not perfectly synchronized.

PN codes are generated using *linear-feedback shift registers* (LSRs) that produce *maximal length* sequences, that is, the longest sequences possible given the number of stages in the register. A maximal-length register with n stages produces codes that are $2^n - 1$ bits long. Phase shifts are generated by multiplying a PN sequence by another bit sequence, called a *PN mask.*

IS-95 uses two types of PN codes:

- *Short PN codes*, $2^{15} - 1$ chips long. At the typical rate of 1.2288 Mc/s, a short code repeats every 26.7 ms.
- *Long PN code*, $2^{42} - 1$ chips long. At the typical rate of 1.2288 Mc/s, a long PN code repeats once every 41.4 days.

UTRAN and cdma2000 use similar codes.

PN codes are used to separate a base station's transmission from those of adjacent cells (which typically use the same carrier frequency) and to separate transmissions from MSs in the same cell. They may also be used to create individual code channels in lieu of Walsh codes. In IS-95 and cdma2000, all cells use the same codes, with phase offsets providing the differentiation; in UTRAN different cells use different codes.

13.1.3 CDMA Channels

Communication between a mobile station and a base station takes place in unidirectional *channels*, separated from one another by chipping codes and by time-division or frequency-division multiplexing. CDMA uses the same, occasionally confusing, nomenclature used by AMPS, TDMA, and GSM (Chapter 12), where the term "channel" is used to denote both an individual channel and a group of channels.

Communication in the forward direction from a BS to all MSs in the cell takes place in the *Forward CDMA Channel* (or *Forward CDMA Link*). Communication in the reverse direction, comprising all transmissions from all MSs in a cell to the BS, takes place in the *Reverse CDMA Channel* (or Reverse *CDMA Link*). The

Forward and Reverse CDMA Channels typically use separate RF bands and are a collection of *code channels* separated by Walsh or PN codes.

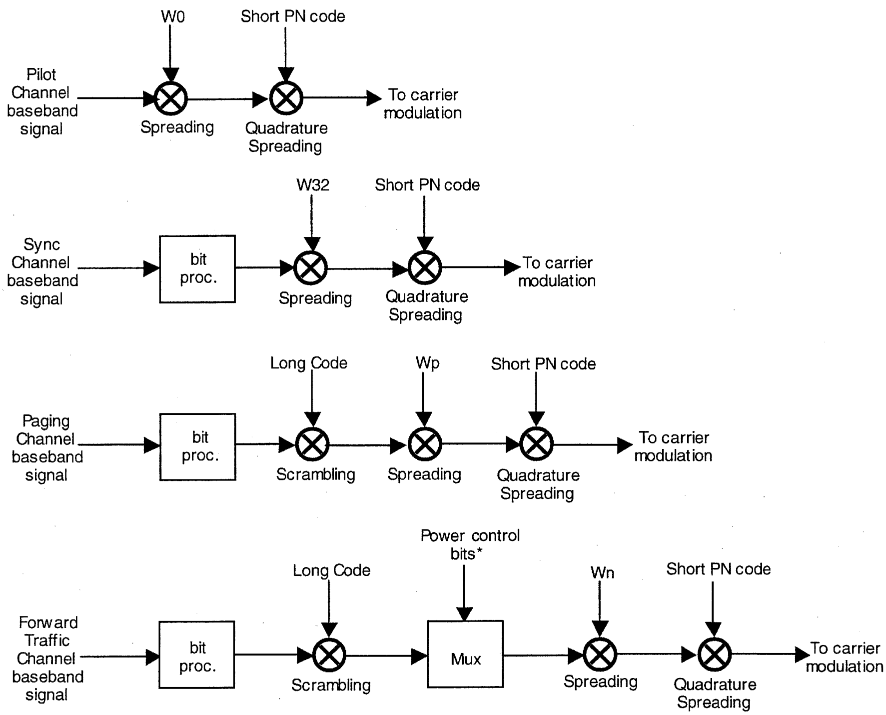
A group of code channels that carry user information to or from a specific MS (a subset of the Forward or Reverse CDMA Channel) constitutes a *Traffic Channel.*

Code channels are categorized as common, shared, or dedicated. Common channels are used by all MSs in a cell. Shared channels are used by a few MSs in a cell based on assignment from the BS. Dedicated channels are MS-specific.

Subchannels may be carved out of code channels by stealing bits or via time slots.

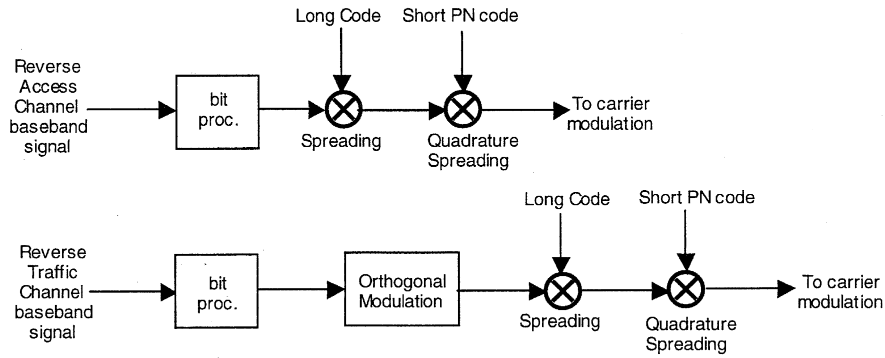
Multistep Spreading and Despreading. The application of chipping codes in CDMA systems is done in steps, with each step performing a particular function and using a different code. Figures 13.2-1 and 13.2-2 show examples of the coding steps at the source in the forward and reverse channel for IS-95 systems. The decoding steps (not shown) at the destination are executed in reverse order.

Only the first coding at the source and the last decoding at the receiver spread and despread the bandwidth of the user signal (e.g., from 14 kHz to 1.22 MHz and vice versa). The other coding and decoding steps, although still called spreading and



W0: Walsh code 0 Wp: Walsh code 1 to 7 Wn: Walsh code 1 to 31 and 33 to 63
bit proc.: bit processing such as convolutional encoding, interleaving, and symbol repetition.
*: only applies to Fundamental Channels

Figure 13.2-1. Conceptual view of Forward CDMA Channel bit processing for IS-95 B.



bit proc.: bit processing such as convolutional encoding, interleaving, and symbol repetition.

Figure 13.2-2. Conceptual view of Reverse CDMA Channel bit processing for IS-95 B.

despreading in most of the literature, do not change the bandwidth because their chip rates are the same as those of the first and last steps.

Before the initial coding step at the source, the user's data stream is subjected to processes that increase the reliability of transmission as well as the bit rate, such as *convolutional encoding* (a form of error correction coding), *symbol repetition* (also used in AMPS, see Section 12.2.2), and *bit interleaving* (in which bits are shuffled to mitigate the effects of error bursts). At the destination the processes are applied in reverse after the final decoding step.

13.1.4 Other Aspects of CDMA

Traffic. We distinguish two types of traffic in CDMA channels: *user traffic* (or *bearer connections*) and s*ignaling traffic* (or *control connections*).

User traffic consists of voice and/or data flows between users. User traffic is subdivided into:

- *Primary traffic*—typically the voice connection
- *Secondary traffic*—additional user data flows, typically nonvoice such as packet traffic

Signaling traffic consists of message flows that are used to manage radio resources and establish, manage, and release connections.

Primary, secondary, and signaling traffic flows can be time-multiplexed in the same code channel with two techniques: *blank and burst* and *dim and burst.*

With blank and burst, an individual frame (see below) is *blanked* for primary traffic and used to send a burst of data with secondary or signaling traffic. More on blank and burst is found in Section 12.2.2.

With dim and burst, some bits in a frame are stolen from primary traffic and used to send signaling and/or secondary traffic, with little effect on voice quality.

Frames. Each code channel consists of a sequence of *frames*, which constitute the basic timing interval for functions such as power control and bit rate control. Frame structures vary by protocol and within each protocol they vary by type of channel and by service configuration. Frames may be subdivided into *subframes*.

Power Control. One of the most challenging problems that had to be solved for the commercialization of CDMA is power control in the reverse CDMA channel. The effect of spreading on received signals is similar to the effect of noise, so it is critical that signals from all MSs arrive at the BS with approximately the same power level. Without power control, strong signals from nearby MSs would crowd out weaker signals from MSs farther away, in what is known as the *near–far effect*. To solve that problem, the power emitted by each MS must be adjusted as the MS moves, to ensure its signal arrives at the BS with constant power. Two methods are used for power control: *open loop* and *closed loop*.

Open Loop Power Control. The sender adjusts its power based on the power of the signal it receives, without any direct feedback from the receiver. For example, a MS adjusts its emitted power based on the power of the signal received from the BS. This method provides a fast reaction time but is not very accurate, because there is no exact symmetry between the two directions of transmission.

Closed Loop Power Control. The entity receiving a signal sends power control information (bits) to the sender, directing it to increase/decrease signal power, based on the quality of the reception. The interval at which power control bits are sent is called the *power control group*, typically 1.25 ms. This method is slower but more accurate than open loop. Closed loop power control has two components: *inner loop power control* and *outer loop power control*.

The inner loop is a faster mechanism by which the receiver controls the sender's power based on a quick measurement of the received signal. The measurement is a signal-to-noise/interference ratio, called SNR or SIR. The inner loop mechanism raises or lowers the power emitted by the sender to meet a *setpoint* value for the SNR that is adjusted dynamically via an *outer loop*.

The outer loop is a slower mechanism that adjusts the target setpoint of the inner loop. The objective is to achieve a longer-term quality target, based on the *frame error rate* (FER).

The reason for a dynamic inner loop target is that the relationship between SNR and FER varies with MS location and speed, so having a dynamic adjustment, rather than a fixed value based on the worst case, improves performance. The inner loop logic that controls the Reverse CDMA Channel is located in the base transceiver station (BTS); the corresponding outer loop logic is located in the base station controller (BSC).

Although theoretically not necessary (the BS does not move), a form of power control is used also in the forward direction, to minimize the power emitted by the BS. The control logic for Forward CDMA Channel power is located in the MS.

Reuse Factor and Soft Handoff. An important advantage of CDMA over TDMA and FDMA is that there is no need to use different RF bands in adjacent cells because separation is achieved through codes. As a consequence, CDMA systems usually have a *frequency reuse factor of one*, compared with the typical factor of seven in AMPS, TDMA, and GSM (see Fig. 12.1-5). That results in a dramatic improvement in bandwidth efficiency, easily gauged by observing that a forward or a reverse (narrowband) CDMA channel requires approximately the same bandwidth and carries approximately the same traffic as a corresponding AMPS or TDMA channel. Because adjacent CDMA cells reuse the same RF band, the bandwidth needed for one cell is also the bandwidth needed for a whole CDMA system. AMPS and TDMA systems, on the other hand, require up to seven times the bandwidth needed by one cell because adjacent cells must use different RF bands, in the typical seven-cell pattern. The actual efficiency gain is between five and seven.

Other benefits of the reuse factor of one are simplified frequency planning and *soft handoffs* [5].

CDMA frequency planning is simpler because a new cell can be added without impacting the RF assignment of adjacent cells.

A soft handoff takes place when a MS is in a border area between two BSs and detects a new cell. The MS starts transmitting to the new cell without dropping the connection to the previous cell; the MSC uses the audio signal from both cells until the signal from the old cell falls below a threshold. At that point the mobile switching center (MSC) disconnects the connection to the old cell and the handoff is complete. The technique (*make before break*) avoids a *hard handoff*, which uses a *break before make* technique where the connection to the old BS is dropped first and then the connection to the new BS is activated. MSs must do a hard handoff when moving into a cell that uses a different carrier frequency.

Registration. Registration is the process by which a MS informs the BS about its location, status, and identity. Registration allows more efficient paging by informing the system of a MS's location, thereby limiting the scope of pages (see also Section 12.3.2). Registration can be initiated by the MS or the BS and can take place when a MS changes location (moves to a new cell), when parameters stored in a MS are changed, and at predetermined time intervals. Call originations and call terminations provide the BS with information equivalent to a registration (implied registration).

13.2 IS-95 AIR INTERFACE

The IS-95 standard supports dual-mode operation: CDMA and analog (AMPS). Analog operation is discussed in Chapter 12; this chapter discusses the CDMA operation, based on the ANSI/TIA/EIA-95B standard [7].

13.2.1 Overview

The network model for IS-95 is essentially the same as the one for AMPS (Section 12.1). Base stations consist of two functional entities—the *base transceiver station* (BTS) and the *base station controller* (BSC). Mobile switching centers (MSCs) provide switching services for the nonaccess (non-air-interface) part of connections.

All base stations in an IS-95 system are synchronized using the global positioning system (GPS), and their common timing reference is called *system time*. The global time reference is needed by BSs and MSs to synchronize their short and long PN code generators so they can produce and detect the phase offsets used to separate channels, as explained in detail in the next sections.

In North America, IS-95 CDMA can be deployed in the 800-MHz frequency band, called the *cellular band*, and in the 1900-MHz band, called the *personal communications system* (PCS) *band*. In either case, a forward or reverse CDMA channel occupies a 1.23-MHz RF band, resulting from a spreading rate of 1.2288 Mc/s. The separation between the Forward and the Reverse CDMA Channel is 45 MHz in the cellular band and 80 MHz in the PCS band. If there is more than one forward and/or reverse CDMA channel in a cell, each occupies a different RF bands.

An IS-95 mobile station supports two bit rates from the vocoder:

- Rate Set I has a maximum input rate of 8 kb/s, which becomes 9.6 kb/s after adding overhead.
- Rate Set II has a maximum input rate of 13.3 kb/s, becoming 14.4 after adding overhead.

Three types of chipping codes are used in IS-95 CDMA systems:

1. Walsh codes
2. Short PN codes
3. Long PN code

Walsh codes are used to separate channels in the forward direction. In the reverse direction they are used to encode the baseband signal, rather than for separating channels. IS-95 systems use the set of 64-chip Walsh codes, named W0 through W63. W0 is all zeros; the other codes have 32 zeros and 32 ones. Cycle time is 0.052 ms.

Short PN codes are used to separate cells or cell sectors from each other. Each BS uses the same pair of short codes and achieves separation from other BSs by using one out of 512 possible phase offsets, each 64 chips apart. The number of chips and cycle time are as discussed in Section 13.1.2.

The Long PN code is used in the Forward CDMA Channel to scramble the baseband signal and in the Reverse CDMA Channel to separate transmissions from individual MSs (via phase offsets). The PN code is the same in all cells. Number of chips and cycle time are as discussed in Section 13.1.2.

Power Control. The Reverse CDMA Channel uses open loop power control for initial channel acquisition and then closed loop power control for traffic channels, with inner and outer loop operations.

Forward CDMA Channel power control is based on a simplified form of closed loop: the MS sends a FER measurement for the received signal to the BS (in a Power Measurement Report message on the reverse traffic channel), which uses it to adjust its power.

13.2.2 Forward CDMA Channel

The Forward CDMA Channel, transmitted over the cell's carrier frequency, is separated from other cells by a specific offset of the short PN codes. A BS may also transmit more than one Forward CDMA Channel by means of frequency-division multiplexing (different RF carriers). A Forward CDMA Channel is subdivided into code channels, separated by Walsh codes. There are four types of code channels; the first three are common channels that serve all the MSs in the cell, while the fourth is dedicated to individual MSs:

1. Pilot Channel
2. Synch Channel
3. Paging Channel
4. Traffic Channel

Walsh code orthogonality can be exploited in the Forward CDMA Channel because MSs are phase-synchronized with the BS. Synchronization is acquired and maintained with the help of the Pilot and Sync Channels. There are 64 possible code channels in the Forward CDMA Channel, based on the number of available Walsh codes. A conceptual view of the bit manipulation process that generates the physical Forward CDMA Channel is shown in Fig. 13.2-1. Each code channel, after being spread with a Walsh code, is spread *in quadrature* with two short PN codes, before modulating the carrier for transmission. Quadrature spreading means that the bit stream is split into two branches, each spread separately with one of two PN sequences, and then modulated by two carriers with the same frequency but 90 degrees out of phase. The two branches, called I (in-phase) and Q (quadrature), are added together before transmission, resulting in quadrature phase-shift keying (QPSK) modulation.

Pilot Channel. This is a common channel used by the mobile station to measure the strength of the signal transmitted by a BS, to select the strongest signal, and to acquire initial synchronization with a BS via short PN code phase-offset detection. There is always one and only one Pilot Channel per Forward CDMA Channel. It carries an unmodulated signal of all zeros and is spread by Walsh code W0. This channel is sent with higher power than the other channels, so it has a high signal-to-noise ratio (SNR), making it easy to acquire.

Sync Channel. This is a common channel used by the BS to transmit (periodically) the Sync Channel message, which contains system information used by mobile stations to complete the synchronization process with the BS. There is a maximum of one Sync Channel per Forward CDMA Channel, spread by Walsh code W32.

Paging Channel. This is a common channel used by the BS to contact individual MSs (*paging*) and to send system information, such as traffic channel assignment. Up to seven code channels can be assigned as paging channels, which can be configured to operate at 4800 or 9600 bps. They are spread by Walsh codes W1 through W7. Channel 1 (primary paging channel) is the default channel.

Paging Channel transmission is structured as a sequence of 20-ms frames, grouped into a cycle of 80-ms *slots*. The MS can operate in *nonslotted* as well as *slotted mode*; that is, the MS can listen for pages in all slots or only in a predetermined set of slots, going idle for the rest of the cycle to save power.

Forward Traffic Channel. This is a group (one or more) of code channels from the BS to an individual MS. It carries primary traffic (typically voice) and signaling traffic and may also carry secondary traffic (additional user data streams). It consist of one forward fundamental channel and up to seven optional forward supplemental code channels, each structured as a sequence of 20-ms frames.

Multiple types of traffic can be multiplexed in the same code channel using blank and burst or dim and burst techniques. The initial bits of each frame are used to define the allocation of the remaining bits to primary, secondary, and signaling traffic (see example of Fig. 13.2-3). The multiplexing of traffic types in a traffic channel and their bit rates are configurable according to sixteen *multiplex options*. Supplemental channel assignment and multiplex options are negotiated between BS and MS during connection setup, by an exchange of messages (*service negotiation* messages in Tables 13.2-2 and 13.2-4).

The bit stream in the fundamental channel is *punctured* with a power control bit every 1.25 ms, or 800 times per second, to form a power control subchannel. A bit value of 1 causes the MS to decrease power by 1 dB; a value of 0 causes it to increases power by 1 dB.

Walsh codes W1 to W63 can be assigned to fundamental and supplemental traffic channels, except for W1–W7 and W32 when used for Paging Channels and the Sync Channel, respectively.

13.2.3 Reverse CDMA Channel

The Reverse CDMA Channel is subdivided into code channels, separated by long PN code phase offsets and structured as a sequence of 20-ms frames. There are

MM	TT	TM	Traffic (168 bits)

MM: Mixed Mode (1 bit) TT: Traffic Type (1 bit) TM: Traffic Mode (2 bits)

Figure 13.2-3. Example of frame format with multiplexed traffic (Multiplex Option 1).

two types of code channels:

1. Reverse Access Channel
2. Reverse Traffic Channel

There is no Pilot Channel in the reverse direction, to reduce MS power consumption. Without a Pilot Channel it is not possible for the MS to acquire and maintain the strict phase synchronization with the BS that would enable the use of Walsh codes for channel separation. Instead, each code channel is spread by the long PN code and separated from the others by a phase offset. After the application of the long code, the signal is spread in quadrature by the same short PN codes used in the Forward CDMA Channel, but with zero offset. A conceptual view of the bit manipulation process that generates the physical Reverse CDMA Channel is shown in Fig. 13.2-2.

The MS sends *preamble bursts* in reverse code channels to aid the BS in acquiring the signal.

Reverse Access Channel. This is a common channel used by MSs for signaling communication with the BS when they have not been assigned a reverse traffic channel. A MS uses the Reverse Access Channel for registration, to respond to page messages, and to originate calls. Once a MS is assigned a traffic channel, signaling communication moves to the Reverse Traffic Channel (see below).

To contact the BS on this channel, the MS uses a random access protocol consisting of a series of *access probes* and listens for a response on the Paging Channel.

An Access Channel is always associated with a Paging Channel and a Paging Channel supports from 1 to 32 Access Channels. The long code phase offset that identifies an Access Channel is applied by means of a *long code mask* derived from base station ID, paging channel number, access channel number, and the short PN code offset of the forward channel.

Reverse Access Channel frames are grouped into a cycle of *slots*, with the number of frames in a slot configurable on a cell basis.

Reverse Traffic Channel. This is a group (one or more) of dedicated code channels that carry primary traffic, secondary traffic, and signaling from an individual MS to the BS. It consists of one Reverse Fundamental Traffic Channel and up to seven optional Reverse Supplemental Code Channels. The same multiplexing options are supported and the same considerations apply as for the Forward Traffic Channel.

The Reverse Traffic Channel uses long PN code phase offsets for channel separation. Each code channel (fundamental and supplemental) in the same traffic channel uses the same long code but with a different phase offset. Long code phase offsets are applied by means of masks derived from the MS's electronic serial number (ESN) and the code channel index (0 for the fundamental and 1–7 for the supplemental channels).

8 bits	(variable length)	16 or 30 bits	
Message Length	Message Body	CRC	Padding

Figure 13.2-4. Signaling message capsule.

Walsh codes are used for modulation, instead of spreading. Every group of six bits from the source, after adding overhead but before spreading with the long code, is used as a pointer (64 possible values) to one of the 64 Walsh codes, which are then used to code the bit stream in a process called *orthogonal modulation.*

13.2.4 Signaling Messages

Signaling messages span multiple frames and have the general format of Fig. 13.2-4. The first three information elements (IEs)—message length, message body, and CRC—constitute the actual message; *padding* is then added, as needed, to fill the last frame of the message. A message with padding constitutes a *message capsule*.

Sync Channel Message. This is the only message sent on the Sync Channel. The message body contains information elements that convey system information to the MS, such as:

- SID: System ID—identifies a region and a service provider
- NID: Network ID
- PILOT_PN: The short PN code phase offset that identifies the BS
- LC_STATE: Long code state (state of the 42 shift register cells in the long code generator)
- SYS_TIME: System time—GPS time in 80-ms units
- LTM_OFF: The offset of local time from system time in 30-min units
- PRAT: Paging channel bit rate
- CDMA_FREQ: Center frequency of the cell's RF carrier

Paging Channel Messages. Examples of Paging Channel message types are listed in Table 13.2-1.

Forward Traffic Channel Messages. The Forward Traffic Channel uses blank and burst and dim and burst to send signaling messages. Examples of message types sent on the Forward Traffic Channel are listed in Table 13.2-2.

Reverse Access Channel Messages. Transmission of a message on the Access Channel starts with a preamble, which is a series of frames containing all zeros, followed by a message capsule. Examples of message types sent on the Reverse Access Channel are listed in Table 13.2-3.

TABLE 13.2-1 Paging Channel Messages

Message Type	Function
Access Parameters	Parameters used by the Reverse Access Channel
Authentication Challenge	Authentication
CDMA Channel List	Available nearby CDMA carriers
Channel Assignment	Directs the MS to a Traffic Channel
Data Burst	Short mesage service, broadcast
General Page	Paging
Neighbor List	Accessible nearby BSs
Order	See Table 13.2-5
SSD Update	Authentication
Status Request	Registration
System Parameters	Overhead information
TMSI Assignment	Privacy

TABLE 13.2-2 Forward Traffic Channel Messages

Message Type	Function
Authentication Challenge	Authentication
Alert With Information	Alerting
General Handoff Direction	Handoff
MS Registered	Registration
Neighbor List Update	Accessible nearby BSs
Order	See Table 13.2-5
Power-Up Function (PUF)	Power-up of MS
Service Request	Service negotiation
Service Response	Service negotiation
Status Request	Registration

Reverse Traffic Channel Messages. Signaling messages are sent using dim and burst or blank and burst. Examples of message types sent on the reverse traffic channel are listed in Table 13.2-4.

Order Messages. An Order message is a type of message that can be sent both on the forward channels (Paging and Traffic) and on the reverse channels (Access and Traffic). Order messages perform different functions based on the value of the order qualification code (ORDQ) and other IEs. Examples of Order message functions are listed in Table 13.2-5.

TABLE 13.2-3 Reverse Access Channel Messages

Message Type	Function
Authentication Challenge Response	Authentication
Order	See Table 13.2-5
Origination	Service negotiation
Page Response	Response to page
Registration	Registration
TMSI Assignment Completion	Privacy

TABLE 13.2-4 Reverse Traffic Channel Messages

Message Type	Function
Authentication Challenge	Authentication
Data Burst	Short message service, broadcast
Handoff Completion	Handoff
Origination Continuation	Call origination
Order	See Table 13.2-5
Periodic Pilot Strength Measurement	Handoff
Pilot Strength Measurement	Handoff
Power Measurement	Power control
Send Burst DTMF	Feature control
Service Connect Completion	Service negotiation
Service Option Control	Service negotiation
Service Request	Service negotiation
Service Response	Service negotiation
TMSI Assignment Completion	Privacy

13.2.5 IS-95 Call Sequences

The IS-95 specification describes call sequences in terms of *call states* for the mobile station,

- MS initialization state
- MS idle state (monitor Paging Channel)
- MS system access state
- MS control on the Traffic Channel state

and in terms of *call processing types* for the base station,

- Pilot and Sync Channel processing (interacts with the MS initialization state)
- Paging Channel processing (interacts with the MS idle and MS system access states)

TABLE 13.2-5 Order Messages

Order Message Function	Forward Channels[a]	Reverse Channels[a]
Base station acknowledgment	P T	
Connect		T
Continuous DTMS tone	T	T
Mobile station acknowledgment		A T
Parameter update	T	
Parameter update confirmation		T
Registration accepted	P	
Registration request	P	
Release	P T	A T
Service option request	T	T
Service option response	T	T
Shared secret data update confirmation		A T

[a]A, Reverse access channel; P, paging channel; T, traffic channel.

- Access Channel processing (interacts with the MS system access state)
- Traffic Channel processing (interacts with the MS control on the Traffic Channel state)

MS Initialization. When a mobile station powers up, it tunes to the CDMA carrier frequency and attempts to acquire the Pilot Channel. Acquisition of the Pilot Channel is relatively simple, even though the MS is not synchronized with the BS yet, because the pilot uses a known Walsh code (W0, consisting of all zeros) and transmits an all-zero baseband pattern. The Pilot is also spread by the short PN code pairs, which are applied with a phase offset specific to the particular BS. The MS despreads the received signal by "sliding" its internally generated short PN codes until the all-zero signal is detected. If pilots are detected at multiple offsets (e.g., when pilots from multiple BSs are detectable in a border area), the MS chooses the one with the strongest signal.

At this point the MS knows that its internally generated short PN sequence is in-phase with the sequence received from the BS. Because the number of chips of the short code is a 512 multiple of a Walsh cycle, the BS can now synchronize its internally generated Walsh codes with the codes received from the BS and decode reverse code channels. The MS then despreads the Sync Channel by applying code W32, and receives the Sync Channel message sent periodically by the BS. The message contains system information (long code state and system time) that enables the MS to synchronize its long code generator with the BS. With short, long, and Walsh codes now synchronized, the MS enters the idle state and monitors paging channel 1. From the System Parameters message, broadcast periodically on the paging channel, the MS learns the number of paging channels supported by the BS. If that number is greater than one, the MS selects one of the supported paging channels with a *hash* algorithm that uses the mobile's ID and the number of channels as inputs.

After the expiration of a timeout started upon entering the idle state, the MS performs a *power-up registration*, informing the BS of its presence in the cell. The BS is then able to select the appropriate Paging Channel for paging the MS, using the same parameters and algorithm as the MS.

Access Probes. Access probes are transmissions used by the MS to establish contact with the BS. There are three situations in which a MS sends probes:

- To register with the BS
- To respond to a paging message
- To originate a call

A MS sends access probes on the Reverse Access Channel and listens for acknowledgment on the Paging Channel. If the BS supports more than one access channel, the MS learns the number of supported channels from the Access Parameters message broadcast by the BS on the Paging Channel. The MS then chooses one of the access channels using a pseudorandom algorithm on the number of access channels.

An access probe begins with a preamble burst (a sequence of all zero frames), immediately followed by a message. Closely spaced probes make an *access probe sequence*, and one or more (15 maximum) access probe sequences make an *access subattempt*, several of which constitute an *access attempt.* Access probes in each sequence are performed with progressively higher power until the BS responds.

MS Registration. Registration takes place when the MS is in the idle state. The registration process starts with an access attempt for a Registration message on the Reverse Access Channel. The BS acknowledges the attempt with a Registration Accepted Order message on the Paging Channel. The BS can also request registration by sending the MS a Registration Request Order or a Status Request message, both on the Paging Channel.

Through registration the BS learns the mobile station's ID.

MS Origination. To originate a call, an idle MS makes an access attempt for an Origination message (with the mobile's ID) on the Reverse Access Channel and listens for acknowledgment on the Paging Channel. If the attempt is successful, the BS assigns a Forward Traffic Channel to the MS (via a Channel Assignment message), on which it sends null traffic data; it then prepares to receive on the Reverse Traffic Channel by generating the appropriate long code phase shift based on the MS ID. The MS configures itself to receive on the assigned Forward Traffic Channel and sends a preamble on the Reverse Traffic Channel. The BS uses the preamble to synchronize with the MS and acknowledges acquisition of the preamble with a Base Station Acknowledgment Order on the Forward Traffic Channel. The MS then starts sending null traffic data on the Reverse Traffic Channel, after which the BS signals the MS to connect and start the conversation.

MS Termination. An example of a call terminated to a MS is shown in Fig. 13.2-5. The MS monitors the Paging Channel and, when the BS sends a page on the Paging Channel, the MS responds with an Access Attempt on the Reverse Access Channel, containing a Page Response message. Upon detection of the Page Response, the BS configures a Forward Traffic Channel, on which it sends null traffic data, and prepares to receive on the Reverse Traffic Channel. The BS also informs the MS of the Forward Traffic Channel assignment via a message on the Paging Channel. At this point all signaling moves to the Traffic Channels. The MS configures itself to receive on the Forward Traffic Channel and starts sending an all-zero preamble on the Reverse Traffic Channel. The BS acknowledges the acquisition of the preamble with a Base Station Acknowledgment Order, upon receipt of which the MS starts sending null traffic data on the Reverse Traffic Channel. The BS then instructs the MS to establish the connection. After an acknowledgment by the MS, the BS instructs the MS to ring; when the called

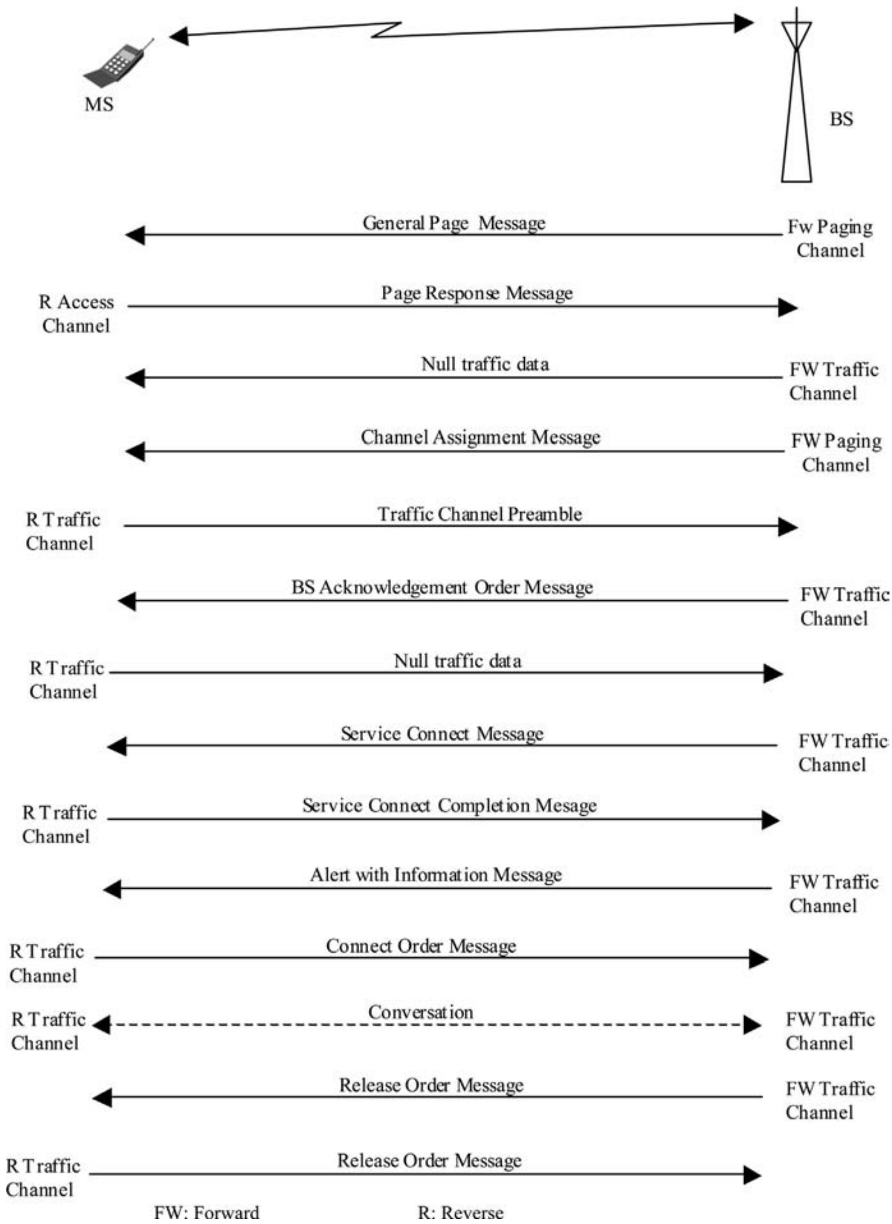


Figure 13.2-5. Call flow example of call terminated to the MS.

party answers, the MS stops ringing and sends a Connect Order message, which starts the conversation. The call can be released either by the MS or by the BS (as in the example), with the sending of a Release Order message, followed by an acknowledgment.

Handoff. Three types of handoff are supported:

- *Soft Handoff.* This type of handoff is possible only if the new cell has the same frequency assignment as the old cell.
- *CDMA-to-CDMA Hard Handoff.* This is used between disjointed cells that use different carrier frequencies.
- *Handoff to Analog.* This is a type of hard handoff and its support is optional for the BS but mandatory for the MS.

Handoffs can take place in the following MS states: idle, system access, and control on the Traffic Channel.

For soft handoffs, the MS monitors pilots from adjacent cells and when it detects a pilot with signal strength exceeding a threshold, it informs the BS with a Pilot Strength Measurement message. The BS uses the information to direct the MS to perform the handoff. An example of a soft handoff message sequence is shown in Fig. 13.2-6.

13.3 cdma2000 AIR INTERFACE

cdma2000, one of the air interface standards recognized by ITU-T under IMT-2000, is the evolutionary offspring of IS-95 CDMA with which it can be made backward

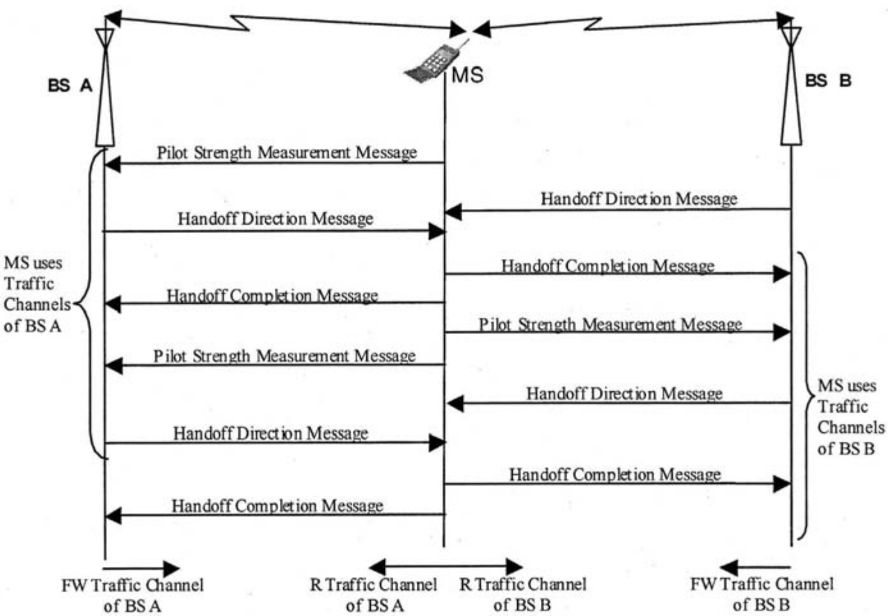


Figure 13.2-6. Call flow example of a soft handoff.

compatible: when properly configured, cdma2000 cells can work with IS-95 mobile stations and cdma2000 mobile stations can operate in IS-95 cells. The main enhancement over IS-95 is the ability to support wideband data concurrently with voice [2,3,8–11].

Like IS-95 CDMA, cdma2000 supports dual-mode operation: CDMA and analog (AMPS—Chapter 12). This chapter describes the CDMA operation [12,13].

cdma2000 is much more complex and flexible than IS-95, with a large number of channels, many possible configurations, and an expanded ability to use multiple channels in parallel to transmit user data.

The cdma2000 specifications are issued jointly by the 3GPP2 Council and by ANSI/TIA/EIA (Section 2.2).

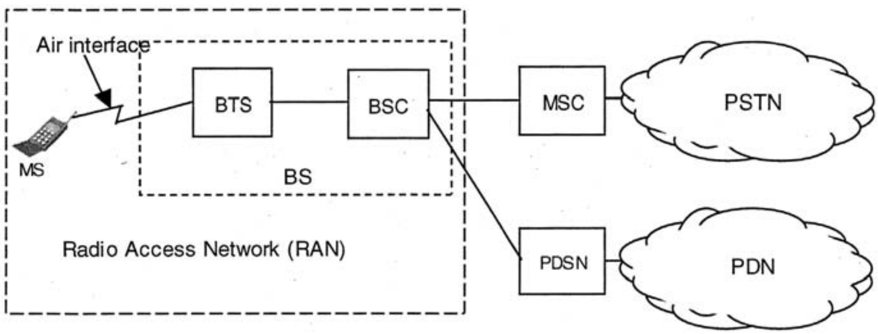
13.3.1 Overview

The network model for cdma2000 is almost the same as the one for AMPS and IS-95 (Fig. 13.3-1) [14]. The main elements are the base station (BS), comprised of a base transceiver station (BTS) and a base station controller (BSC), the mobile switching centers (MSCs), and the *packet data service node* (PDSN). The PDSN connects the *radio access network* (RAN) to the packet data network, to support wideband data communication. A description of the PDSN can be found in [15,16].

Like IS-95, cdma2000 systems are synchronized via GPS.

Spreading Rates (SRs). cdma2000 supports two spreading modes: *spreading rate 1* (SR-1) and *spreading rate 3* (SR-3).

SR-1 uses one RF carrier per direction, each with a 1.23-MHz bandwidth and a 1.2288-Mc/s chip rate. SR-1 service is called 1X and has the same chip rate and bandwidth as IS-95 systems but provides faster packet data rates concurrent with voice.



PDN: Packet Data Network

Figure 13.3-1. cdma2000® network and interfaces.

SR-3, also called multi-channel RF (CDMA-MC) mode, has the following characteristics:

- The Forward CDMA Channel uses three separate carriers, each with a 1.23-MHz bandwidth and a 1.2288-Mc/s chip rate.
- The Reverse CDMA Channel uses one 3.69-MHz bandwidth carrier with a 3.6864-Mc/s chip rate.

SR-3 service is called 3X and supports true wideband data rates.

Radio Configurations. Traffic Channels can be configured according to different *radio configurations* (RCs). A RC is a package that defines which code channels can work together and the physical layer characteristics of each channel such as data rates, chip rates, Walsh code lengths, and other configurable parameters. Ten radio configurations (RC 1 through RC 10) are supported for the Forward Traffic Channel, and seven (RC 1 through RC 7) for the Reverse Traffic Channel. Two radio configurations (RC 1 and RC 2 in either direction) are backward compatible with IS-95. Tables 13.3-1, 13.3-2, 13.3-4, and 13.3-5 show the channel combinations allowed by each radio configuration.

Data Rates. Three types of bit rates are supported for Traffic Channels: fixed, flexible, and variable.

Fixed rates are predefined for a given radio configuration.

Flexible rates are configurable, typically in 1-bit increments, under the control of the BS and apply to RC 3 and above. The BS sends a flexible rate table (in the Service Connect, General Handoff Direction, or Universal Handoff Direction messages) that defines the number of bits per frame in a channel.

Variable rate can vary, within a limited set, on a frame-by-frame basis. They apply to traffic channels in all radio configurations, except to R-SCCH and F-SCCH (defined later), which have fixed rates.

Variable rate operation is supported within flexible rate operation.

Channel Structure. cdma2000 uses a much larger number of channels than IS-95 and distinguishes between *physical channels* and *logical channels*.

Physical channels are connections between MS and BS that are defined by the way the information is transferred at layer 1 of the protocol, that is, by physical characteristics such as chipping codes and power control method.

Logical channels are logical connections between peer entities in the MS and BS and are defined by the type of information they carry. They are controlled by upper layer applications. Multiple logical channels can be mapped into the same physical channel, based on information in the protocol data unit (PDU) headers (Section 13.3.4).

Figure 13.3-2 shows how logical and physical channels relate to the protocol stack. Physical channels map into logical channels at the MAC sublayer of the stack.

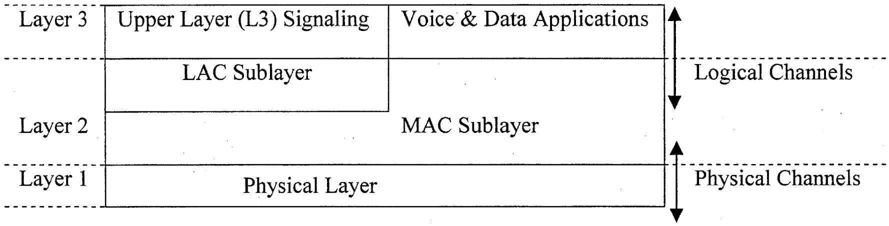


Figure 13.3-2. Protocol stack and physical and logical channels for cdma2000®.

Physical channels are always denoted by uppercase acronyms, while logical channels use lowercase acronyms. The acronyms are based on simple rules: F/f for forward, R/r for reverse, D/d for dedicated, C/c for common, CH/ch for channel, t for traffic and s for signaling. The type and number of channels actually used in a system, out of all the ones defined in the standard, depend on radio configurations and spreading rates.

Channels, both physical and logical, can be common or dedicated. A common channel is shared by all MSs in a cell and is used to send and receive information to/from MSs before they have been assigned a traffic channel, for example, during initialization, idle, and system access states.

Dedicated channels are used by individual MSs to exchange signaling and user data with the BS, and they are assigned by the BS. The BS can assign multiple dedicated channels to the same MS to form a *Traffic Channel.*

A significant difference with IS-95 is found in the Reverse CDMA Channel, which can be configured (in RC 3 and higher) with Reverse Pilot Channels. The synchronization made possible by Reverse Pilot Channels allows the use of Walsh codes in the reverse direction.

Power Control. In the Reverse CDMA Channel, power control with RC 1 and RC 2 works the same way as for IS-95. With RC 3 through RC 7 the operation is almost the same, with open loop and closed loop, but there are some differences in the way the power control subchannel is derived. A Common Power Control Channel (see next section) may also be used.

In the Forward Channel, power control with RC 1 and RC 2 works exactly as for IS-95, while RC 3 through RC 10 use a full closed loop operation with inner and outer loop, both controlled by the MS via the Reverse Power Control Subchannel.

13.3.2 cdma2000 Signaling

Protocol Stack. The communication protocol for the cdma2000 air interface is structured with three layers and several sublayers (Fig. 13.3-2):

- Layer 1, or physical layer
- Layer 2, or data link layer, subdivided into two sublayers: medium access control (MAC) and link access control (LAC)
- Layer 3, or upper layer signaling (ULS)

The LAC sublayer is present only in the portion of the protocol stack used for signaling and is absent from the user data portion.

Due to the complexity of the protocol, the following sections describe message components at a level higher than bit level. Readers interested in the bit level of detail are referred to [17–24].

Primitives. cdma2000 standard documents describe internal communication between adjacent layers and sublayers (in the BS or in the MS) in terms of *primitives.* Primitives contain the parameters needed for communication, in particular, the SDUs to be exchanged and the channel or channel type to be used.

13.3.3 Physical Layer

The *physical layer* manages the physical aspects of communication between the MS and the BS, such as frequencies, timing, modulation, demodulation, coding, spreading, despreading, and power control [17,18].

Two chip rates are supported: 1.2288 Mc/s and 3.6864 Mc/s. With SR-1, 1.2288 Mc/s is used both in the forward and in the reverse channels. With SR-3, 1.2288 Mc/s is used in the forward channel, and 3.6864 Mc/s is used in the reverse channel.

The physical layer interfaces with the MAC sublayer on one side and with the peer physical layer at the far end of the connection on the other side. Communication takes place via *physical channels*, created through a combination of Walsh codes, long and short PN codes, and time slots. Physical channels come in a variety of frame structures (some configurable) and support different bit rates.

Tables 13.3-1 and 13.3-2 list the physical channels and their groupings.

Chipping Codes. The same three types of chipping codes used in IS-95 are used in cdma2000, namely, Walsh codes, short PN codes, and the long PN code, with two major enhancements:

1. Walsh codes in both the Forward and the Reverse CDMA Channel
2. Variable-length Walsh codes

Walsh codes are used in the Reverse CDMA Channel with RC 3 and higher configurations, allowing a larger number of code channels to be supported compared to IS-95. The use of Walsh codes requires the presence of Reverse Pilot Channels.

In both the Reverse and the Forward CDMA Channel, Walsh codes are of variable length, based on channel type and radio configuration. The different lengths support multiple data rates: longer codes are used for lower rates and shorter codes for higher rates (the total chip rate is constant). The Reverse CDMA Channel uses Walsh codes with length of 2, 4, 8, 16, and 64 chips. The Forward CDMA Channel uses Walsh codes with length of 8, 16, 32, 64, 128, and 256 chips.

TABLE 13.3-1 Forward CDMA Physical Channels

Physical Channels	Acronym	Grouping	Maximum Number of Channels		RC
			SR-1[a]	SR-3[b]	
Auxiliary Pilot Channels	F-APICH	Pilot Channels	U/1	U/1	—
Auxiliary Transmit Diversity Pilot Channel	F-ATDPICH	Pilot Channels	U/1	NA	—
F Pilot Channel	F-PICH	Pilot Channels	1/1	1/1	—
Transmit Diversity Pilot Channel	F-TDPICH	Pilot Channels	1/1	NA	—
Common Power Control (sub)Channel	F-CPCCH	F Indicator Control Channel	U/1	4/1	—
F Rate Control (sub)Channel	F-RCCH	F Indicator Control Channel	U/1	NA	—
F Dedicated Control Channel	F-DCCH	F Traffic Channels	U/1	U/1	3–9[c]
F Fundamental Channel	F-FCH	F Traffic Channels	U/1	U/1	1–9
F Packet Data Channel	F-PDCH	F Traffic Channels	2/1	NA	10
F Power Control Subchannel	NA	F Traffic Channels	[d]	[d]	[d]
F Supplemental Channel	F-SCH	F Traffic Channels	U/2	U/2	3–9[e]
F Supplemental Code Channel	F-SCCH	F Traffic Channels	U/7	NA	1–2
Broadcast Control Channel	F-BCCH	NA	8/1	8/1	—
Common Assignment Channel	F-CACH	NA	7/1	7/1	—
F Acknowledgment Channel	F-ACKCH	NA	1/1	NA	—
F Common Control Channel	F-CCCH	NA	7/1	7/1	—
F Grant Channels	F-GCH	NA	U/2	NA	—
F Packet Data Control Channel	F-PDCCH	NA	2/2	NA	10
Paging Channel	F-PCH	NA	7/1	NA	—
Quick Paging Channel	F-QPCH	NA	3/1	3/1	—
Sync Channel	F-SYNCH	NA	1/1	1/1	—

[a]SR-1: Maximum number of channels per BS/maximum number of channels monitored by one MS for SR-1.
[b]SR-3: Maximum number of channels per BS/maximum number of channels monitored by one MS for SR-3.
[c]For SR-1 only RC 3–5, for SR-3 only RC 6–9.
[d]Part of the Forward Fundamental Channel or of the Forward Dedicated Control Channel.
[e]For SR-1 only, RC 3–5.

Note: F, forward; RC, radio configuration; SR, spreading rate; NA, not applicable; U, undefined.

TABLE 13.3-2 Reverse CDMA Physical Channels

			Maximum Number of Channels per MS		
Physical Channels	Acronym	Grouping	SR-1	SR-3	RC
Access Channel	R-ACH	Legacy Access Operation	1	NA	—
Enhanced Access Channel	R-EACH	Enhanced Access Channel Operation	1	1	—
R Pilot Channel	R-PICH	Enh. Acc. Ch Op, RCC Ch Op	1	1	3–7
R Common Control Channel	R-CCCH	RCC Channel Operation	1	1	3–7
R Acknowledgment Channel	R-ACKCH	R Traffic Channels	1	NA	3–7
R Ch Quality Indicator Channel	R-CQICH	R Traffic Channels	1	NA	3–7
R Dedicated Control Channel	R-DCCH	R Traffic Channels	1	1	3–6
R Fundamental Channel	R-FCH	R Traffic Channels	1	1	1–6
R Packet Data Channel	R-PDCH	R Traffic Channels	1	NA	7
R Packet Data Control Channel	R-PDCCH	R Traffic Channels	1	NA	7
R Power Control Subchannel	[a]	R Traffic Channels	[b]	[b]	3–7
R Request Channel	R-REQCH	R Traffic Channels	1	NA	7
R Secondary Pilot Channel	R-SPICH	R Traffic Channels	1	NA	7
R Supplemental Channel	R-SCH	R Traffic Channels	2[a]	2	3–6
R Supplemental Code Channel	R-SCCH	R Traffic Channels	7	NA	1,2

[a]RC 3 and RC 4 only.

[b]Part of the Reverse Pilot Channel.

Note: R, reverse; RC, radio configuration; RCC, reverse common control; Enh., enhanced; Op, operation; Acc., access; Ch, channel; NA, not applicable (channel type available for SR-1 only).

For forward channels Walsh codes are assigned by the BS. For reverse channels, Walsh codes have fixed values that depend on channel type. Every MS uses the same codes because transmissions from individual MSs are separated from each other by long code phase offsets.

The short PN codes and the long PN code used with SR-1 are the same as for IS-95 (Section 13.1.2).

For SR-3, the short PN codes in the forward direction are the same as for SR-1; in the reverse direction they are based on a sequence $2^{20} - 1$ chips long, truncated to a length of 3×2^{15} chips. The long PN code in the forward direction is the same as for SR-1; in the reverse direction it is derived from the standard SR-1 code generator with a multiplexing operation that boosts it up to the higher chip rate.

Forward CDMA Channel. Channel separation is provided by Walsh codes. With RC 1 and RC 2, spreading is the same as for IS-95 (Fig. 13.2-1).

With RC 3 and above, after Walsh spreading, channels are spread by a *complex multiplier* with the two short PN codes, whose phase offset provides BS identification. Most channels, such as Traffic Channels, are also scrambled with the long code before spreading. Others, such as the Pilot and Sync Channels, are not. A high-level conceptual view of the bit manipulation process that generates the physical Forward CDMA Channel is shown in Figure 13.3-3(*a*).

Reverse CDMA Channel. In RC 1 and RC 2, the Access Channel and the Traffic Channels are fully compatible with IS-95: they use Walsh codes for orthogonal modulation and long PN code offsets for channel separation (Fig. 13.2-2).

In RC 3 and above, each common channel and each group of dedicated channels transmitted by a given MS is separated by a long code phase shift, generated by a mask that depends on the channel type:

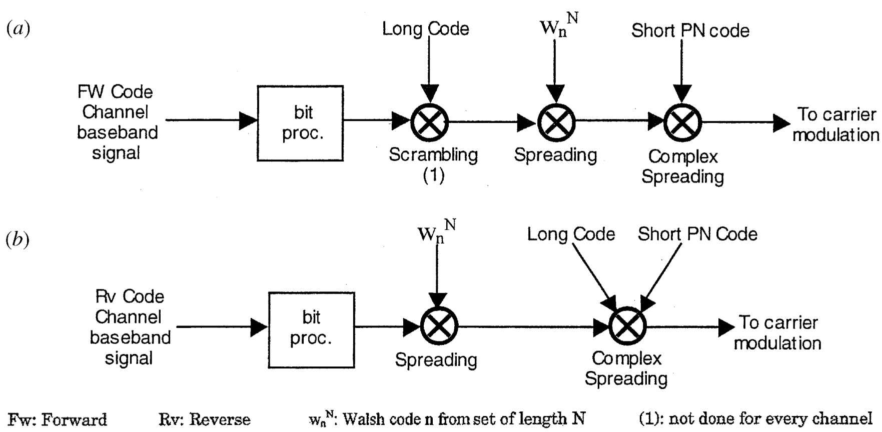


Figure 13.3-3. Conceptual view of physical channel bit processing for cdma2000® (RC 3 and above).

- For common channels, all MSs use common masks that depend on channel type. To avoid collisions, a multiple access protocol is used, and the address of the transmitting MS is sent to the BS in the signaling communication.
- For dedicated channels, each MS uses a MS-specific mask, based on ESN or IMSI (see Section 13.3.6).

The dedicated code channels from an individual MS are separated by Walsh codes, which require the presence of a reverse pilot. Each MS transmits a Reverse Pilot Channel as one of the code channels. After orthogonal spreading with Walsh codes, the long PN code sequence is applied, together with the short PN sequences, by a *complex multiplier*. A high-level conceptual view of the bit manipulation process that generates the physical Reverse CDMA Channel is shown in Fig. 13.3-3(*b*).

Physical Channels. Compared to IS-95, cdma2000 supports a much higher number of physical channels and provides much greater flexibility in their configuration. Channel configuration (which group of channels performs what function) is negotiated between the BS and the MS via layer 3 signaling messages such as Origination and Channel Assignment (Sections 13.3.6 and 13.3.7). Below are brief descriptions of the physical channels.

F-FCH: Forward Fundamental Channel. This dedicated channel carries user traffic (voice and data) as well as signaling and power control. For data traffic it supports data rates of up to 14.4 kb/s. When higher data rates are needed, Supplemental Channels must be added to the connection. A Fundamental Channel can time-multiplex user and control data but, to avoid possible performance impact, a separate Dedicated Control Channel may be configured for signaling. The frame can be 5 or 20 ms.

F-SCH: Forward Supplemental Channel. This dedicated channel is used optionally, in conjunction with F-FCH, for user-data rates higher than those supported by F-FCH alone. Up to two F-SCHs can be allocated to a MS, each with a different rate. Maximum data rate is 1.0368 Mb/s with SR-3. The frame can be 20, 40, or 80 ms. It is supported only in RC 3 through RC 9.

F-SCCH: Forward Supplemental Code Channel. This dedicated channel is used optionally, in conjunction with F-FCH, for user-data rates higher than those supported by F-FCH alone. Up to seven F-SCCHs may be assigned to a MS. The maximum data rate is 14.4 kb/s and the frame is 20 ms. It is supported only in RC 1 and RC 2 (SR-1) and is backward compatible with IS-95.

F-PICH: Forward Pilot Channel. This common channel is used by the MS as phase reference for timing acquisition (short PN code phase offset) as well as a reference for measurements needed for handoffs.

F-APICH: Auxiliary Pilot Channel. This forward pilot channel is used for spot beams (antenna emissions over a narrow angle) and beam forming (smart antenna technique that generates a narrow targeted beam).

F-SYNC: Sync Channel. This common channel is used by the BS to broadcast system information and by a MS to acquire full synchronization once it has acquired a F-PICH. The frame is 26.7 ms.

F-BCCH: Broadcast Control Channel. This common channel is used to broadcast control information and overhead data to MSs when they have not been assigned a traffic channel. It is paired with F-CCCH and used instead of F-PCH when IS-95 compatibility is not required. Up to eight F-BCCHs are supported in a Forward CDMA Channel: the first is called the Primary F-BCCH and is used for overhead data; the others, when present, are called Secondary F-BCCHs and are used for broadcasts. The frame is 40 ms.

F-QPCH: Quick Paging Channel. This channel is used optionally to inform a MS that a paging message is expected on the Paging Channel or the Forward Common Control Channel. It allows the MS to save power by ignoring paging unless necessary.

F-PCH: Paging Channel. This common channel is used by the BS to page a MS and also to send periodic overhead messages (Access Parameters message). It is backward compatible with IS-95. The frame is 20 ms.

F-CACH: Common Assignment Channel. This common channel is used to respond to access requests sent on the Enhanced Access Channel. The frame is 5 ms.

F-DCCH: Forward Dedicated Control Channel. This dedicated channel is used in RC 3 through RC 9 to send signaling and power control information. It may also be used for sending bursts of data, replacing dim and burst and blank and burst operations. The frame can be 5 or 20 ms.

Forward Power Control Subchannel. This subchannel is time-multiplexed in F-FCH and F-DCCH to control the power of the MS when it has been assigned a Traffic Channel.

F-TDPICH: Transmit Diversity Pilot Channel. This pilot channel is used to support transmit diversity, a forward transmission technique that uses multiple antennas. It works in conjunction with F-PICH.

F-ATDPICH: Auxiliary Transmit Diversity Pilot Channel. This pilot channel supports transmit diversity (see above) and works in conjunction with F-APICH.

F-CCCH: Forward Common Control Channel. This common channel is used to send control information to MSs when they have not been assigned a Traffic Channel. When compatibility with IS-95 systems is not necessary, it can be used for paging instead of F-PCH. In that case it is paired with F-BCCH, which handles overhead and broadcasts. When this channel is used instead of F-PCH, the MS responds to pages on the R-EACH instead of the R-ACH. The frame can be 5, 10, or 20 ms.

F-CPCCH: Common Power Control Channel. This channel is used for power control when a F-FCH or a F-DCCH is not assigned, and when a R-CCCH is used. It is divided into subchannels, each assigned to a different MS. This channel is sometimes referred to as being part of the Forward Indicator Control Channel: for SR-3, the F-CPCCH is the only component, in which case the two are one and the same; for SR-1, F-CPCCH shares the Forward Indicator Control Channel with F-RCCH. The frame is 10 ms.

F-PDCH: Forward Packet Data Channel. This shared channel is used to carry packet (user) data. A BS may support up to two of these channels. Data for multiple MSs are time-multiplexed into one channel, under control of the F-PDCCH. It is used only with SR-1 and RC 10. The operation of the F-PDCH involves the F-PDCCH, the R-CQICH, and the R-ACKCH. The frame can be 1.25, 2.5, or 5 ms.

F-PDCCH: Forward Packet Data Control Channel. This channel is used to send control information to MSs for the operation of the F-PDCH. Messages on this channel contain the ID of the destination MS. Up to two channels may be supported by a BS. It is used only with SR-1 and RC 10. The frame can be 1.25, 2.5, or 5 ms.

F-ACKCH: Forward Acknowledgment Channel. This feedback channel is used to send acknowledgments of the reception of packet transmissions on the R-PDCH. The frame is 10 ms.

F-GCH: Forward Grant Channel. This channel is used to send information on the packet size to be used by the MS on the R-PDCH. The frame is 10 ms.

F-RCCH: Forward Rate Control Channel. This channel is used to control the data rates of the R-PDCH. It is divided into subchannels, each of which may control one or more R-PDCHs. This channel is sometimes referred to as part of the Forward Indicator Control Channel, which it shares with F-CPCCH. The frame is 10 ms.

R-FCH: Reverse Fundamental Channel. This dedicated channel carries user data with a maximum rate of 14.4 kb/s. For higher data rates, Supplemental Channels must be added to the connection. Signaling may be time-multiplexed with user data on the R-FCH or may be sent separately on the Dedicated Control Channel (R-DCCH). The frame can be 5 or 20 ms.

R-SCH: Reverse Supplemental Channel. This dedicated channel is used, in conjunction with R-FCH, to support user-data rates higher than what can be handled by the R-FCH alone. Up to two optional R-SCH channels can be assigned to a MS. The maximum data rate is 1.0368 Mbps with SR-3. The frame can be 20, 40, or 80 ms. It is supported only in RC 3 through RC 6.

R-SCCH: Reverse Supplemental Code Channel. This dedicated channel is used, in conjunction with R-FCH, to support data rates higher than what can be handled by the R-FCH alone. Up to seven R-SCCHs are allowed for one MS. The maximum data rate is 14.4 kb/s. The frame is 20 ms. It is supported only in RC 1 and RC 2 (SR-1) and is backward compatible with IS-95.

R-ACH: Access Channel. This channel is used to request a connection to the BS, to respond to pages, and for registration. It is used only with RC 1 and RC 2 (SR-1) and is compatible with IS-95. Other configurations, not IS-95 compatible, use the R-EACH and the R-CCCH for the same functions. The frame is 20 ms.

R-EACH: Enhanced Access Channel. This channel is used for access when paging is done on F-CCCH and is not backward compatible with IS-95. It operates according to two modes: *basic access mode* and *reservation access mode*. In the latter case, the access procedure also involves F-CACH and R-CCCH. The frame can be 5, 10, or 20 ms.

R-PICH: Reverse Pilot Channel. This channel is used by the BS for phase reference, allowing the use of Walsh codes in the reverse direction. It transmits all zeros, is spread by Walsh code 0 of length 64, and can carry power control bits, time-multiplexed with the phase reference. It is used in RC 3 and above.

Reverse Power Control Subchannel. This subchannel can be inserted, under certain conditions, into R-PICH to control BS power. It is divided into primary and secondary subchannels. It is supported only in RC 3 to RC 7.

R-DCCH: Reverse Dedicated Control Channel. This dedicated channel is used to send user and control (signaling) data to the BS. The frame can be 5 or 20 ms. It is supported only in RC 3 through RC 6.

R-CCCH: Reverse Common Control Channel. This common channel is used to send control information to the BS when the MS has not been assigned a Traffic Channel. Typical use is when the MS accesses the network on the R-EACH in reservation access mode. It is power-controlled by the F-CPCCH. The frame can be 5, 10, or 20 ms.

R-PDCH: Reverse Packet Data Channel. This shared channel is used for packet data (user) transmission. The operation of the R-PDCH involves the R-PDCCH, the

R-REQCH, the F-ACKCH, the F-GCH, the F-RCCH, and the R-SPICH. The frame is 10 ms. It is supported only for RC 7 and SR-1.

R-PDCCH: Reverse Packet Data Control Channel. This channel is used to send control information for the operation of the R-PDCH. The frame is 10 ms.

R-ACKCH: Reverse Acknowledgment Channel. This channel is used to send acknowledgments of the reception of packet transmissions on the F-PDCH. The frame is 1.25 ms.

R-CQICH: Reverse Channel Quality Indicator Channel. This channel is used to send quality transmission measurements and the selection of the serving sector (by the MS) for the operation of the F-PDCH. The frame is 1.25 ms.

R-REQCH: Reverse Request Channel. This channel is used to send buffer status and "power headroom" information for the operation of the R-PDCH. The frame is 10 ms.

R-SPICH: Reverse Secondary Pilot Channel. This pilot channel is sent, for 10-ms intervals under control of the MAC sublayer, to assist the BS in the reception of some R-PDCH transmissions.

Frames, Slots, and Data Rates. Physical channel frames come in 5-, 10-, 20-ms formats. Some channels have fixed-length frames but most are configurable. Frame configuration is negotiated between the BS and MS (with layer 3 signaling) prior to the establishment of a connection, such as during paging or at call origination; it can also be renegotiated during the connection. The vast majority of physical channels have the frame format shown in Fig. 13.3-4. The encoder tail is a fixed bit sequence used by the channel encoder to reset itself at the end of the frame.

In some channels, a given number of consecutive frames are grouped into *slots*. For instance, F-PICH and F-CCCH transmissions are divided into 80-ms slots. A MS may listen to a channel only in certain slots, to save power (slotted mode).

Bit rates are also configurable, and supported rates vary from channel to channel based on radio configuration, frame size, and spreading rate. Table 13.3-3 gives an indication of the ranges of possible rates.

Data	ET	CRC

ET: Encoder Tail CRC: Cyclic Redundancy Check

Figure 13.3-4. cdma2000® Frame.

TABLE 13.3-3 cdma2000 Data Rates for Bearer Channels (in bps)

	SR-1		SR-3	
Channel	Lowest	Highest	Lowest	Highest
R-FCH	1,200	14,400	1,500	14,400
R-SCH	1,200	307,200	1,200	1,036,800
R-PDCH	19,200	1,845,600	NA	NA
F-FCH	1,200	14,400	1,500	14,400
F-SCH	1,200	307,200	1,200	1,036,800
F-PDCH	81,600	3,091,200	NA	NA

Radio Configurations. Traffic channels in the physical layer are configurable according to radio configurations (Section 13.3.1). Supported configurations are summarized in Tables 13.3-4 and 13.3-5.

In both the Forward and the Reverse Traffic Channel, RC 1 and RC 2 are fully compatible with IS-95 and correspond to Rate Set 1 and Rate Set 2, respectively. The physical characteristics of the traffic channels in those two configurations are the same as those of the corresponding IS-95 channels. All other configurations use Reverse Pilot Channels and Walsh codes in the Reverse CDMA Channel.

Multiple radio configurations can be used concurrently. However, a MS using RC 1 or RC 2 on the Reverse CDMA Channel cannot use any of the other radio configurations, because of the obvious incompatibility. For the same reason, if a BS uses RC 1 or RC 2 on a Forward Traffic Channel, it cannot use any of the other radio configurations on that channel.

TABLE 13.3-4 Radio Configurations for the Forward Traffic Channel

RC	Spreading Rate	Channels
RC 1, RC 2	SR-1	F-FCH, F-SCCH
RC 3–RC 5	SR-1, SR-3	F-FCH, F-DCCH, F-SCH
RC 6–RC 9	SR-1, SR-3	F-FCH, F-SCH
RC 10	SR-1	F-PDCH

TABLE 13.3-5 Radio Configurations for the Reverse Traffic Channel

RC	Spreading Rate	Channels
RC 1, RC 2	SR-1	R-FCH, R-SCCH
RC 3, RC 4	SR-1	R-PICH, R-DCCH, R-FCH, R-SCH, R-CQICH, R-ACKCH, R PC Sch.
RC 5, RC 6	SR-3	R-PICH, R-DCCH, R-FCH, R-SCH, R-CQICH, R-ACKCH, R PC Sch.
RC 7	SR-1	R-PICH, R-SPICH, R-PDCH, R-PDCCH, R-REQCH, R-CQICH, R-ACKCH, RPC Sch.

Note: RPC Sch., Reverse Power Control Subchannel.

13.3.4 Medium Access Control (MAC) Sublayer

The MAC sublayer's main function is to manage the mapping and multiplexing of physical channels into logical channels [19,20]. The logical channels supported by the MAC sublayer are:

- Common Signaling Channels (csch): f-csch and r-csch
- Dedicated Signaling Channels (dsch): f-dsch and r-dsch
- Dedicated Traffic Channels (dtch): f-dtch and r-dtch
- Packet Data Channels (pdch): f-pdch and r-pdch
- Forward Broadcast Traffic Channel (f-btch)

At its upper boundary, the MAC sublayer interfaces with the LAC sublayer for signaling traffic, and with upper layers for voice and data traffic (Fig. 13.3-2).

Multiplex and QoS Sublayer. This is the main component of the MAC sublayer and is itself subdivided into three parts (Fig. 13.3-5):

1. Common channel multiplex sublayer (CCMS), which handles common channels

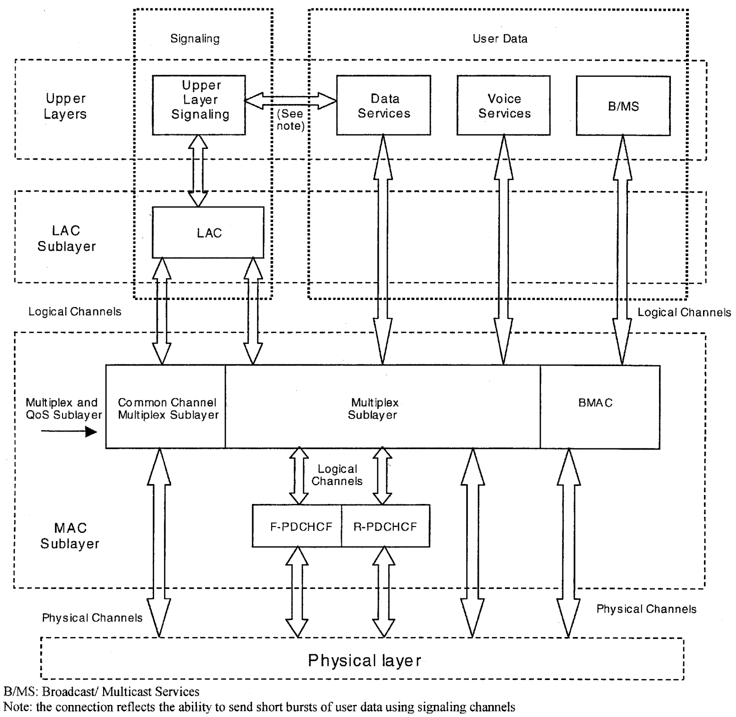


Figure 13.3-5. Detailed protocol structure for the cdma2000® air interface.

2. Multiplex sublayer (MS), which handles dedicated channels
3. Broadcast/multicast multiplex access control (BMAC) sublayer, which handles broadcast and multicast channels

Examples of physical-to-logical channel mapping are shown in Tables 13.3-6 and 13.3-7.

A more detailed depiction of the channel structure in the MAC sublayer is shown in Figs. 13.3-6 and 13.3-7.

The multiplex and QoS sublayer exchanges data with logical channels in the form of *data blocks*. Data blocks from a logical channel or from multiple logical channels (depending on channel type) are mapped into a physical channel. For example, F/R-PDCH and F/R-FCH can carry both signaling and user data, each coming from a different logical channel. The various combinations are controlled by *multiplex options* and a *logical-to-physical mapping* (LPM) table.

In the transmit direction, one or more data blocks are combined, after adding a header, into a *MuxPDU*, whose header determines which components (primary, secondary, and signaling traffic) are included. Figure 13.3-8 shows an example, in which a MM value of zero would indicate that the PDU contains only primary traffic, while a MM value of 1 would indicate that the PDU contains more than one type of traffic, as determined by the values of the TT and TM fields. One or more MuxPDUs are combined to form a MAC SDU. The MAC SDU is sent to the physical layer in a primitive.

TABLE 13.3-6 Examples of Mapping of Physical to Logical Channels

Logical Channel	Physical Channel	MAC Sublayer that Processes the Channel
r-csch	R-ACH, R-EACH	Common channel multiplex sublayer
f-csch	F-PCH, F-BCCH, F-CCCH, F-SYNCH	Common channel multiplex sublayer
f/r-dsch	F/R-DCCH, F/R-FCH, F/R-SCH*i*, F/R-PDCH	Multiplex sublayer
f/r-dtch	F/R-DCCH, F/R-FCH, F/R-SCH*i*, F/R-PDCH	Multiplex sublayer
f-btch	F-SCH (note)	BMAC multiplex sublayer

Note: Multiple instances of this type of channel may be present on the same interface.

TABLE 13.3-7 Mapping of Logical Packet Channels into Signaling and Data Logical Channels

Logical Packet Channel	Signaling	Voice and Data
r-pdch	r-dsch	r-dtch
f-pdch	f-dsch	f-dtch

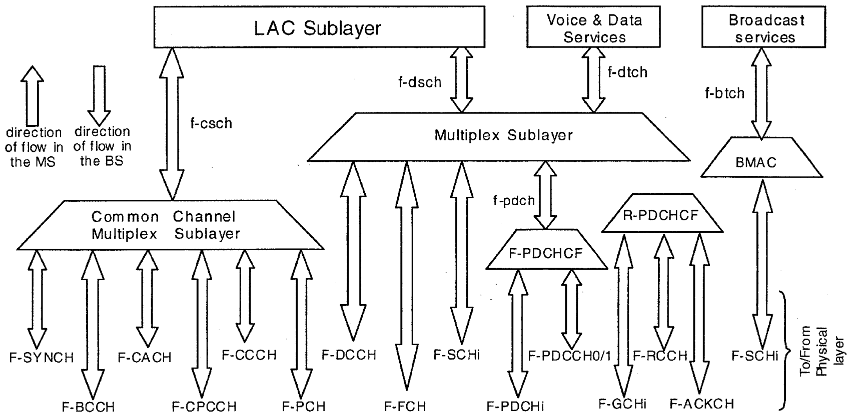


Figure 13.3-6. MAC layer's Forward Channel structure.

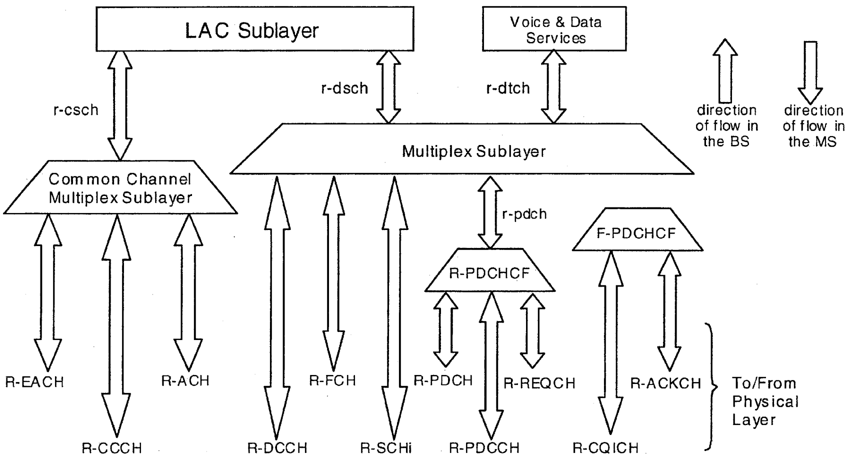


Figure 13.3-7. MAC layer's Reverse Channel structure.

Figure 13.3-9 shows the transmit operation of the multiplex and QoS sublayer for the case where there is no PDCHCF function. The receive operation works in reverse: the sublayer separates the information received in a MAC SDU (from the physical layer) into its components and then routes the components to the appropriate functional entities (as upper layer SDUs) by means of primitives.

Header			Traffic Data
MM	TT	TM	Primary, or Signaling, or Primary & Signaling, or Primary & Secondary

MM: Mixed Mode TT: Traffic Type TM: Traffic Mode

Figure 13.3-8. Example of Mux PDU format for FCH, DCCH, SCCH, SCH, PDCH (Mux PDU Type 1).

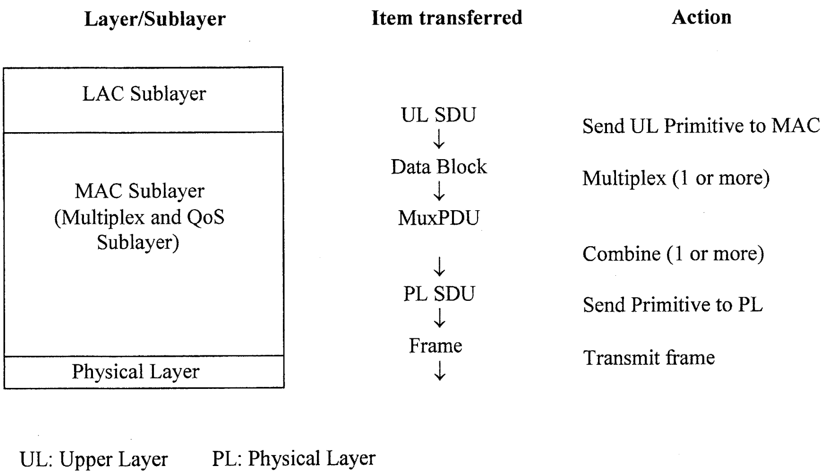


Figure 13.3-9. Transmit operation for the Multiplex and QoS sublayer.

The multiplex and QoS sublayer also determines the relative priority of requests from different sources, in order to ensure quality of service.

F/R-PDCHCF. For communication using the Packet Data Channels, the multiplex sublayer uses the services of the Forward and the Reverse Packet Data Channel control functions (F-PDCHCF, R-PDCHCF). The two functions process PDUs for the Forward and Reverse Packet Data Channel, respectively. How the F/R-PDCHCF relate to the physical and logical channels involved in packet-channel operation is shown in Figs. 13.3-6 and 13.3-7. The figures show that some of the physical channels involved in packet operation are used only in layer 2 communication between peer PDCHCF entities and do not map into logical channels.

The operation of F-PDCH involves the following channels and functions:

- F-PDCH carries user (packet) traffic to the MS.
- F-PDCCH controls the multiplexing of MSs.
- F-CPCCH is used only if the F-PDCH is assigned without a F-FCH or a F-DCCH. In all other cases this channel is not used, since power control is provided by subchannels embedded in F-FCH or F-DCCH.
- R-CQICH sends quality transmission measurements about F-PDCH reception by the MS, which the BS uses to control power and data rates.
- R-ACKCH sends acknowledgments of user-packet reception.

The operation of R-PDCH involves the following channels and functions:

- R-PDCH carries user (packet) traffic from the MS.
- R-PDCCH controls the operation of the R-PDCH.

- R-REQCH sends buffer status and power headroom info for R-PDCH.
- R-PICH assists the BS in the reception of R-PDCH.
- R-SPICH provides further assistance to the BS during reception of certain R-PDCH transmissions.
- F-ACKCH sends acknowledgments of user-packet reception.
- F-GCH sends information on the packet size that the MS can send.
- F-RCCH controls data rates on R-PDCH.

13.3.5 Signaling Link Access Control (LAC) Sublayer

The LAC sublayer is the interface between upper layer signaling and the MAC sublayer [21,22]. The LAC uses logical channels for communication, which insulate it from the physical aspects of the channels. The logical channels used by the LAC are f/r-csch and the f/r-dsch (Figs. 13.3-6 and 13.3-7).

A high-level view of the protocol stack for the LAC is shown in Fig. 13.3-10. The structure of the authentication, integrity, ARQ, and addressing sublayers is dependent on the type of logical channel, and for the details we refer the reader to [21,22]. Each component (sublayer) of LAC adds a header, and in some cases also a trailer, as shown in the examples of Fig. 13.3-11. If a PDU is too long to be sent in one physical-layer frame, it is segmented into *PDU fragments* (Fig. 13.3-12).

The interfaces between the LAC and the MAC sublayer and between the LAC and L3 signaling are called *service access points* (SAPs); the transfer of information at SAPs is via primitives. Primitives exchanged with L3 signaling contain the signaling SDU (i.e., the L3 PDU) and a group of parameters called the *message control and status block* (MCSB), used between internal LAC sublayers and to manage

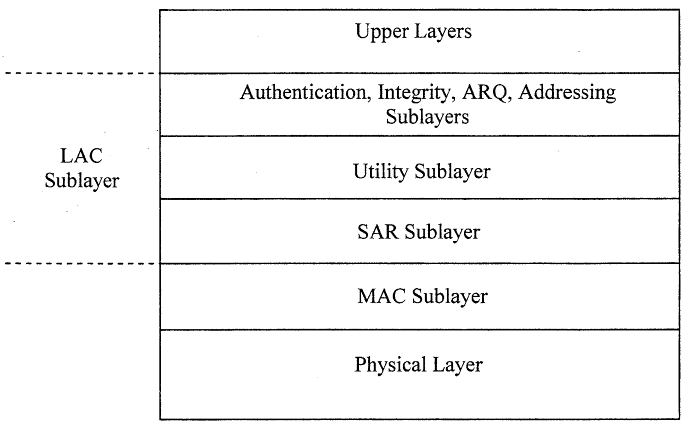


Figure 13.3-10. High-level view of the LAC protocol stack.

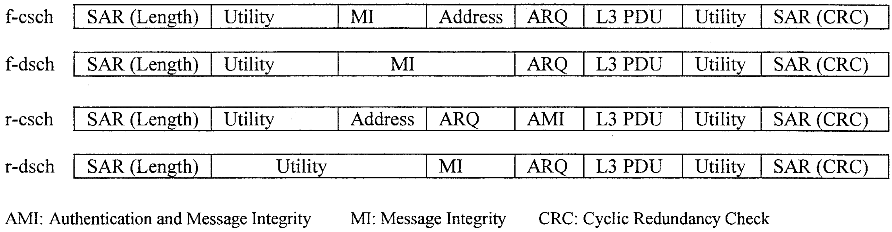


Figure 13.3-11. LAC PDU structure before SAR segmentation.

the layer 3 interface. In the receive direction, the MCSB is generated by the SAR sublayer when it receives a SDU from MAC.

When transmitting, the LAC sublayer builds LAC PDUs, or PDU fragments, and then passes them (as data blocks) to the MAC sublayer. The process is visualized in Fig. 13.3-13. When receiving, the process works in reverse.

Automatic Repeat Request (ARQ) Sublayer. The ARQ sublayer adds acknowledgment fields to the PDU to provide different levels of reliability in message transfer. Reliability is managed by automatic retransmission of messages until an acknowledgment is received.

Two types of ARQ operations are supported: *assured mode* and *unassured mode*. With assured mode a message is resent periodically until an acknowledgment is received. With unassured mode no acknowledgment is sent or expected.

Authentication, Message Integrity, Addressing, and Utility Sublayers. These sublayers add fields to the PDU for authentication, message integrity, MS identification, message type, and padding (the number of octets in a PDU must be an integer). The message type is communicated by layer 3 in the MCSB and is inserted in the header by the utility sublayer as the MSG_TYPE field.

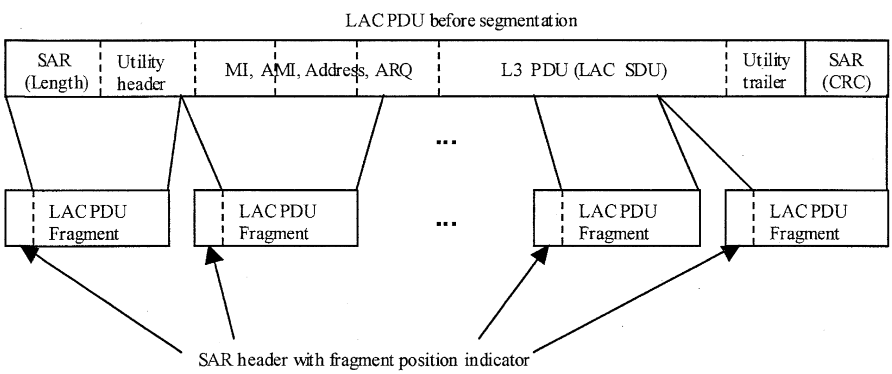


Figure 13.3-12. SAR segmentation of LAC PDUs into encapsulated PDU fragments.

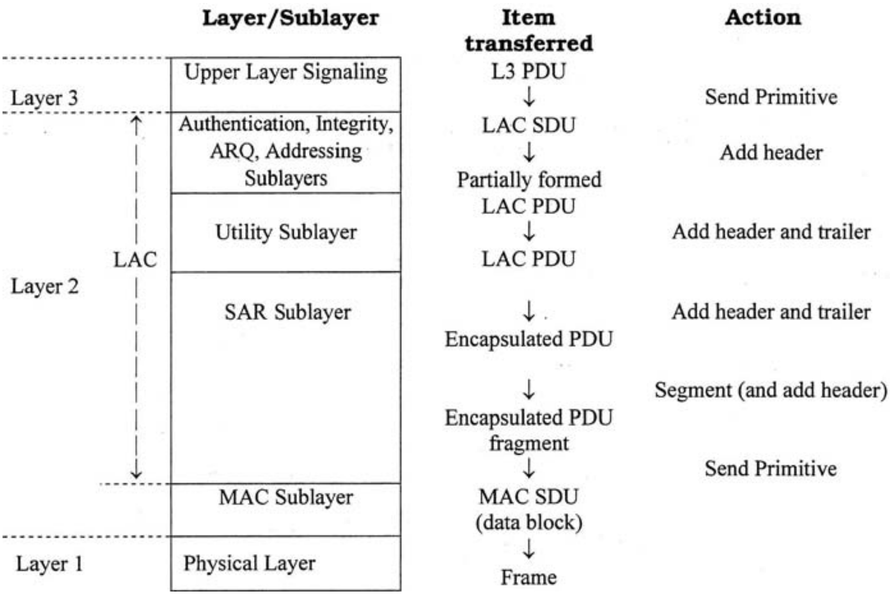


Figure 13.3-13. LAC sublayer—PDU/SDU processing.

SAR Sublayer. On transmission, the segmentation and reassembly (SAR) sublayer breaks the PDU into fragments, based on the frame capacity available from the MAC sublayer. On reception, SAR reassembles the PDU by concatenating successive PDU fragments based on the length of the *PDU capsule* (complete PDU), determined from information in the first fragment.

13.3.6 Upper Layer Signaling (Layer 3)

Upper layer signaling manages the signaling traffic between MS and BS [23,24]. Signaling traffic consists of messages exchanged between a MS and a BS for two basic purposes:

- Manage MS–BS control functions such as power-up, registration, and handoff
- Establish and manage user-to-user connections

To identify and reach a given MS, cdma2000 layer 3 signaling uses several codes and identifiers:

- International mobile subscriber identity (IMSI)
- Mobile directory number (MDN)
- Mobile identification number (MIN)
- Electronic serial number (ESN)

- Mobile equipment identifier (MEID)
- User identity module ID (UIM_ID)
- Shared secret data (SSD)

A MS is primarily identified by the IMSI (see Section 12.7.4). The MS must store only the lower 10 digits of the IMSI, called MIN (see Section 12.3).

The MS is also required to store the MDN, which is a dialable number of up to 15 digits. The MDN is assigned by the service provider and may or may not be the same as the IMSI.

The ESN and MEID codes are assigned by the manufacturer of the MS and the MS needs to store only one of them. The ESN is used with older revisions of the protocol and MEID with more recent ones.

If a removable UIM is present, then UIM_ID is used instead of ESN/MEID. The UIM is a removable module in a MS that carries user information.

SSD is a 128-bit code, not accessible by the user, used for authentication. The authentication procedure verifies that both the BS and the MS have the same SSD (see Section 12.6.7). Authentication and message integrity procedures use ESN, MEID, UIM_ID, and SSD.

Signaling States. The layer 3 signaling process is structured according to four MS states:

1. Initialization state
2. Idle state
3. System access state
4. Mobile station control on the Traffic Channel state

In the *initialization state* the MS powers up, chooses CDMA or analog operation, and connects to and synchronizes with the BS by acquiring the Pilot Channel and then the Sync Channel.

When in the *idle state*, the MS is activated and listens to the f-csch. The f-csch maps into the Quick Paging Channel, or the Paging Channel, or the Forward Common Control Channel and the Primary Broadcast Control Channel, depending on the configuration. In this state the MS is ready to originate a call, to terminate a call, or to start the registration process.

In the *system access state*, the MS exchanges signaling messages with the BS on the common signaling channels (f-csch and r-csch) in order to respond to pages, initiate calls, or initiate the registration process. The r-csch maps into either the Access Channel or the Enhanced Access Channel.

In the *mobile station control on the Traffic Channel state*, the MS is connected with the BS on the traffic channels (f/r-dsch and f/r-dtch), which carry dedicated signaling and user-data traffic. In this state the parameters to be used for traffic channel communication are negotiated, such as radio configurations, frame

TABLE 13.3-8 Upper Layer Signaling (L3) Messages

Message Name	Acronym	f-csch	f-dsch	r-csch	r-dsch
Access Parameters message	APM	X			
Alert With Information message	AWIM		X		
Channel Assignment message	CAM	X			
Extended Channel Assignment message	ECAM	X			
General Page message	GPM	X			
Order message	ORDM	X	X	X	X
Origination message	ORM			X	
Page Response message	PRM			X	
Registration message	RGM			X	
Service Connect Completion message	SCCM				X
Service Connect message	SCM		X		
Service Request message	SRQM		X		X
Service Response message	SRPM		X		X
Status Request message	STRQM	X	X		
Status Response message	STRPM			X	X
Synch Channel message	SCHM	X			
System Parameters message	SPM	X			

structure, data transmission rates, and the types and number of channels (and sub-channels) that constitute the forward and reverse traffic channels for that call.

Messages (L3 PDUs). The most significant layer 3 signaling messages are shown in Table 13.3-8. Order messages perform a variety of functions based on the value of the *order qualification code* (ORDQ) information element. Examples of Order functions are shown in Table 13.3-9.

TABLE 13.3-9 Order Messages

Order Message Function	Forward Logical Channels	Reverse Logical Channels
Base Station Acknowledgment	fdsch, fcsch	NA
Connect	NA	rdsch
Continuous DTMF Tone	fdsch	rdsch
Mobile Station Acknowledgment	NA	rdsch, rcsch
Parameter Update Confirmation	NA	rdsch
Parameter Update	fdsch	NA
Registration Accepted	fcsch	NA
Registration Request	fcsch	NA
Release	fdsch, fcsch	rdsch, rcsch
Service Option Request	fdsch	rdsch
Service Option Response	fdsch	rdsch
SSD Update Confirmation	NA	rdsch, rcsch

Note: NA, not applicable (message not used in this direction).

Information Elements. L3 messages are a sequence of IEs [23,24], which can be optional, nonoptional, or conditional (i.e., present or absent based on the values of other IEs). Due to the large number of IEs, we only list some examples:

- ACC_CHAN: Number of access channels
- BASE_ID: Base station identification
- CDMA_FREQ: Frequency assignment (center frequencies)
- CH_IND: Channel indicator—physical channel allocation
- CHAR: Character—individual dialed digit or character
- CODE_CHAN: Code channel index on the forward channel
- FOR_FCH_RC: Forward Fundamental Channel radio configuration
- LC_STATE: Long code state (phase)
- NID: Network identification
- ORDQ: Order qualification code (indicates the type of Order message)
- PAGE_CHAN: Number of paging channels
- PILOT_PN: Pilot PN (short PN code) offset index
- RECORD_TYPE: Information record type (e.g., registration information)
- RETURN_CAUSE: Reason for the mobile station registration or access
- REV_FCH_RC: Reverse Fundamental Channel radio configuration
- SERV_CON_SEQ: Service connect sequence number
- SERV_REQ_SEQ: Service request sequence number
- SERVICE_OPTION: Requested service option for the call
- SID: System identification
- SYS_TIME: System time
- Information records: Packages of IEs that provide specific information

Examples of service options are basic variable rate voice service, short message service, and packet data service [25,26]. Examples of information records are: service configurations (used in SCM, SRQM, SRPM), calling and called party identity (used in AWIM, ORM), and QoS (used in ORM).

Table 13.3-10 shows how the IE listed above relate to the messages in Table 13.3-8.

13.3.7 cdma2000 Call Sequences

The basic call sequences for initialization, registration, standard (nonpacket) originations/terminations, and handoff are essentially the same as for IS-95 (Section 13.2.5). Differences are mostly concentrated at layers 1 and 2, in which some configurations use multiple physical channels for functions where IS-95 uses one. But for every basic function, a mode of operation that works exactly like IS-95 can always be configured. A typical example is paging and access, described in more detail below.

TABLE 13.3-10 Information Elements in L3 Signaling Messages

	Message Acronyms																
Information Element	APM	AWIM	CAM	ECAM	GPM	ORDM	ORM	PRM	RGM	SCCM	SCHM	SCM	SPM	SRQM	SRPM	STRPM	STRQM
ACC_CHAN	M																
BASE_ID													M				
CDMA_FREQ			C	C							M						
CH_IND							O	O									
CHAR							O										
CODE_CHAN			C	C													
FLC_STATE											M						
NID											M		M				
ORDQ						M											
PAGE_CHAN													M				
PILOT_PN	M		C	C							M		M				
RECORD_TYPE		M										O		O	O	M	M
RETURN_CAUSE							M		M								
REV_FCH_RC				C													
SERV_CON_SEQ										M		M					
SERV_REQ_SEQ														M	M		
SERVICE_OPTION					O		O	M									
SID			C	C							M		M				
SYS_TIME											M						
Information Records		O					O					O		O	O	O	

Note: M, mandatory; O, optional; C, conditional (present or not based on the value of another IE).

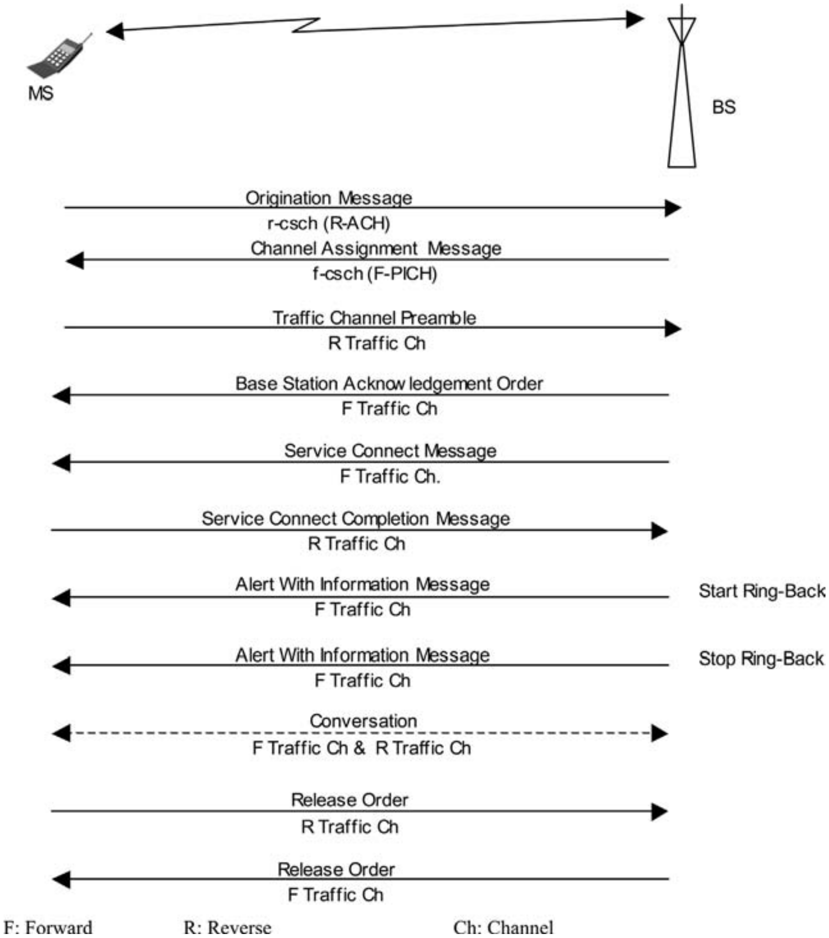


Figure 13.3-14. Example of message flow on the Forward and Reverse Traffic Channels for a MS-originated call.

Figure 13.3-14 shows an example of a call that uses nonpacket physical channels in the traffic channel. Figure 13.3-15 shows an example of a call that uses packet channels, managed by the PDCHCF in the MAC sublayer. In this case some of the messages (denoted as L2) are exchanged between peer PDCHCF functions to initialize the Packet Channels, and L3 signaling is not involved.

Paging and Access Operations. Two types of operation are supported for paging and access:

1. Access Channel/Paging Channel operation, IS-95 compatible
2. Enhanced Access Channel operation, supported only by cdma2000

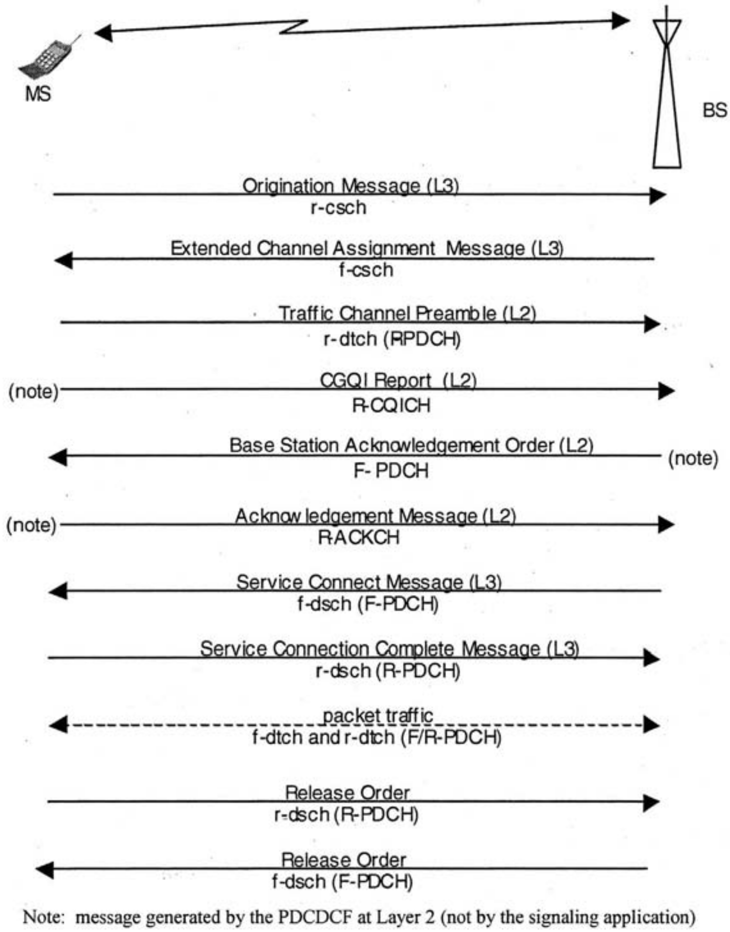


Figure 13.3-15. Example of message flow on the packet channels (F-PDCH and R-PDCH).

The Access Channel/Paging Channel operation is the same as for IS-95: the BS uses the Paging Channel (F-PCH) for paging and the MS uses the Access Channel (R-ACH) for originations, for registrations, and for responding to pages. The F-PCH carries periodic messages with overhead information (System Parameters message, Access Parameters message) and page messages. With this type of operation the Reverse Pilot Channel (R-PICH) is not used.

With the Enhanced Access Channel operation, the BS uses the Primary Broadcast Control Channel (F-BCCH) for periodic messages with overhead information (Enhanced Access Parameters message, ANSI-41 System Parameters message) and the Forward Common Control Channel (F-CCCH) for paging messages. The MS uses the Enhanced Access Channel (R-EACH), the Reverse Pilot Channel

TABLE 13.3-11 Channels for Access and Paging Operations

Physical Channels	Access Channel Operation	Enhanced Access Channel Operation	
		Basic Mode	Reservation Access Mode
F-PCH			
R-ACH			
F-CPCCH			
F-BCCH			
F-CCCH			
R-EACH			
R-PICH			
R-CCCH			

(R-PICH), and, depending on the access mode, the Reverse Common Control Channel (R-CCCH) for originations and for responding to pages. Two modes are supported for the operation of the Enhanced Access Channel:

- Basic mode
- Reservation access mode

In *basic mode* the MS sends an Enhanced Access Channel preamble on the R-PICH, followed by access data on R-EACH and continuous transmission on R-PICH. In this mode the R-CCCH is not used.

In *reservation access mode* the MS sends an Enhanced Access Channel preamble on the R-PICH, followed by an Enhanced Access Channel header on the R-EACH, and then access data on R-CCCH. In this mode the F-CPCCH is used for power control of the R-CCCH.

The physical channel structure for these operations is summarized in Table 13.3-11.

13.4 UTRAN AIR INTERFACE

UTRAN (UMTS Terrestrial Radio Access Network) is the other air interface standard recognized by ITU-T under IMT-2000 [6,9,27–30,33]. UTRAN follows GSM in terms of network architecture and network interfaces but shares with cdma2000 the radio interface technology and the ability to provide wideband data concurrent with voice.

UTRAN is the result of the standardization efforts of the 3GPP council and ETSI (Section 2.2).

13.4.1 Overview

UTRAN Terminology. The UTRAN standards use slightly different terminology from what is commonly used in CDMA literature:

UTRAN	CDMA Literature
Uplink	Reverse
Downlink	Forward
User equipment (UE)	Mobile station (MS)
Radio network subsystem (RNS)	Base station (BS)
Node B	Base transceiver station (BTS)
Radio network controller (RNC)	Base station controller (BSC)
Scrambling	Spreading (when bit rate is unchanged)
Control plane	Signaling message flow
User plane	User-data message flow
Handover	Handoff
Radio frame	Frame
USIM	UIM

The following sections use the common CDMA terminology whenever doing so facilitates understanding.

UMTS Network Model. The UMTS network is divided into two parts (Fig. 13.4-1) [31,32]:

1. Radio access network (RAN)
2. Core network (CN)

The radio access network includes satellite links but this chapter describes only the terrestrial portion of UTRAN. UTRAN covers the portion of the cellular network from the mobile station (UE) to the base station controller (RNC). The reference point between the UE and the base station (RNS) is called Uu, and the one between the base transceiver station (Node B) and RNC is called Iub.

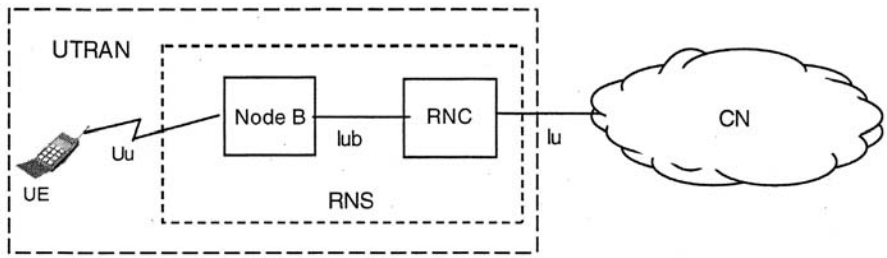


Figure 13.4-1. UMTS network and interfaces.

The UMTS core network evolved from the GSM core network and it includes RNCs and connections to the PSTN, ISDN, and public data networks. The reference point between the RNS and the CN (RNC) is called Iu. The Iu interface is a packet interface, and the UTRAN specifications allow two options: ATM or IP.

The mobile station functionality is subdivided into two parts:

1. Mobile equipment (ME)
2. User services identity module (USIM)

The ME is the basic portion that contains the radio access functions and the logic for the operation of the UE.

The USIM contains the data necessary for identification and security. The USIM is typically implemented as a removable unit.

3GPP documents divide network entities and protocols into an *Access Stratum* (AS) and a *NonAccess Stratum* (NAS). The Access Stratum encompasses the functional entities and protocol layers that manage the radio portion of a wireless network; the NonAccess Stratum covers entities and protocol layers that manage the rest of the network.

UTRAN Air Interface. The UTRAN air interface has two operational modes:

1. Frequency-division duplex (FDD)
2. Time-division duplex (TDD)

With FDD the uplink and downlink radio links use two separate, but paired, frequency bands. With TDD both links share the same frequency band, and radio frames are divided into time slots to create a path in each direction. TDD is intended for situations where there are frequency spectrum constraints; FDD is the standard operational mode and is the one covered in this chapter.

UTRAN base stations are not synchronized, which leads to differences in chipping codes and channel acquisition techniques, compared to IS-95 and cdma2000.

Power Control. Power control uses open loop for initial transmissions, then closed loop operation with inner and outer loop, in both the uplink and downlink direction.

13.4.2 Signaling

The UTRAN air interface is structured along three layers and several sublayers (see Fig. 13.4-2).

In the control plane (signaling) the three layers are:

- Layer 1, or *physical layer*
- Layer 2, subdivided into two sublayers: *medium access control* (MAC) and *radio link control* (RLC)
- Layer 3, or *radio resource control* (RRC)

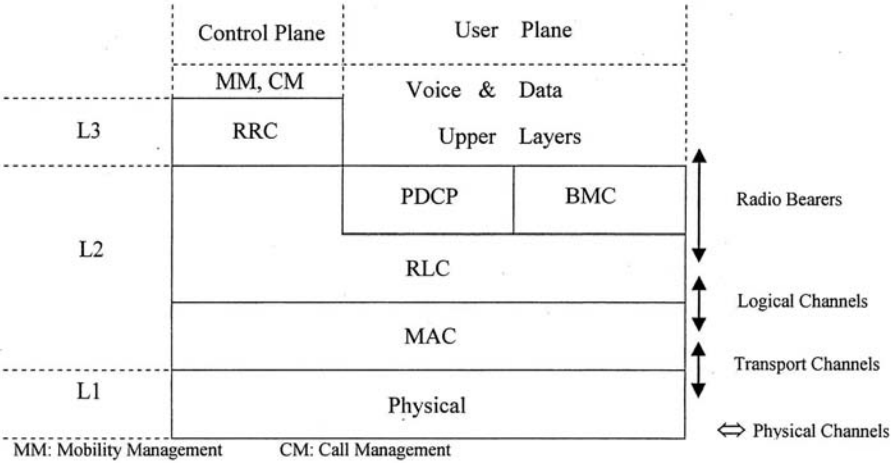


Figure 13.4-2. Protocol stack for the UTRAN air interface.

In the user plane (user-data traffic) the structure of the protocol stack is slightly different: layer 2 has two (parallel) sublayers on top of RLC,

- Packet Data Convergence Protocol (PDCP)
- Broadcast/Multicast Control (BMC)

and layer 3 has no RRC sublayer.

Channel Structure. UTRAN uses channels for communication with peer entities and between layers [33–35]. The channel structure is more complex than the cdma2000 structure and consists of four channel types.

Physical Channels connect peer entities in the physical layers on either side of a radio link, that is, in the MS and the BS. They are identified by frequency and chipping codes (Table 13.4-1).

Transport Channels connect the physical layer to the MAC sublayer of layer 2, a function that in cdma2000 is covered by physical channels (Table 13.4-2).

Logical Channels connect the MAC sublayer to the RLC sublayer within layer 2, and are identified by the type of information they carry (Table 13.4-3).

Radio bearers (RBs) are an abstract way of looking at groups of channels, to insulate the upper layers from the details of the air interface protocol and to allow the core network to interwork with different types of air interfaces (*radio access technologies*—RATs). They are used to manage the transmission and protocol characteristics (e.g., frequency and bit rates) of groups of channels, in order to achieve the QoS level required by the user. In the control plane

TABLE 13.4-1 Physical Channels

Direction	Channel Name	Type	Acronym	Frame
DL	Acquisition Indicator Channel	C	AICH	20 ms, 15 sl
	E-DCH Absolute Grant Channel	C	E-AGCH	10 ms, 15 sl*
	High-Speed Physical Downlink Shared Channel	C	HS-PDSCH	10 ms, 15 sl*
	MBMS Indicator Channel	C	MICH	10 ms, no sl
	Paging Indicator Channel	C	PICH	10 ms, no sl
	Primary Common Control Physical Channel	C	P-CCPCH	10 ms, 15 sl
	Primary Common Pilot Channel	C	P-CPICH	10 ms, 15 sl
	Secondary Common Control Physical Channel(s)	C	S-CCPCH	10 ms, 15 sl
	Secondary Common Pilot Channel(s)	C	S-CPICH	10 ms, 15 sl
	Shared Control Channel	C	HS-SCCH	10 ms, 15 sl*
	Synchronization Channel (Primary and Secondary)	C	(P- and S-) SCH	10 ms, 15 sl
	Dedicated Physical Channel	D	DPCH	10 ms, 15 sl
	E-DCH Hybrid ARQ Indicator Channel	D	E-HICH	10 ms, 15 sl*
	E-DCH Relative Grant Channel	D	E-RGCH	10 ms, 15 sl*
	Fractional Dedicated Physical Channel	D	F-DPCH	10 ms, 15 sl
UL	Dedicated Physical Control Channel	D	DPCCH	10 ms, 15 sl
	Dedicated Physical Data Channel(s)	D	DPDCH	10 ms, 15 sl
	E-DCH Dedicated Physical Control Channel	D	E-DPCCH	10 ms, 15 sl*
	E-DCH Dedicated Physical Data Channel	D	E-DPDCH	10 ms, 15 sl*
	HS Dedicated Physical Control Channel	D	HS-DPCCH	10 ms, 15 sl*
	Physical Random Access Channel	C	PRACH	10 ms, 15 sl

Note: D, dedicated; C, common; DL, downlink; UL, uplink; sl, slots.

*Timing based on 3-slot subframes.

TABLE 13.4-2 Transport Channels

Direction	Channel Name	Subtype	Acronym	Mapping to Physical Channels
DL	Broadcast Channel	C/S	BCH	P-CCPCH, CPICH
	Forward Access Channel	C/S	FACH	S-CCPCH
	High-Speed Downlink Shared Channel	C/S	HS-DSCH	SCH, AICH, PICH, MICH, HS-PDSCH, HS-SCCH, HS-DPCCH
	Paging Channel	C/S	PCH	S-CCPCH
	(DL) Dedicated Channel(s)	D	DCH	DPCH, F-DPCH
UL	Random Access Channel	C/S	RACH	PRACH
	(UL) Dedicated Channel(s)	D	DCH	DPDCH, DPCCH
	Enhanced Dedicated Channel	D	E-DCH	E-DPDCH, E-DPCCH, E-AGCH, E-RGCH, E-HICH

Note: D, dedicated; C, common; S, shared; DL, downlink; UL, uplink.

TABLE 13.4-3 Logical Channels

Direction	Type	Channel Name	Acronym	Mapping to Transport Channels
DL	CCH	Broadcast Control Channel	BCCH	BCH, FACH
		Common Control Channel	CCCH	FACH
		Dedicated Control Channel	DCCH	FACH, HS-DSCH, DCH
		Paging Control Channel	PCCH	PCH
		MBMS Point-to-Multipoint Control Channel	MCCH	FACH
		MBMS Point-to-Multipoint Scheduling Channel	MSCH	FACH
	TCH	Common Traffic Channel	CTCH	FACH
		Dedicated Traffic Channel	DTCH	FACH, HS-DSCH, DCH
		MBMS Point-to-Multipoint Traffic Channel	MTCH	FACH
UL	CCH	Common Control Channel	CCCH	RACH
		Dedicated Control Channel	DCCH	RACH, DCH, E-DCH
	TCH	Dedicated Traffic Channel	DTCH	RACH, DCH, E-DCH

Note: DL, downlink; UL, uplink; CCH, Control Channels; TCH, Traffic Channels.

they connect the RLC sublayer to the RRC sublayer and in the user plane they connect the RLC sublayer to voice and data upper layers.

UTRAN channels are categorized as common, dedicated, and shared. The common (shared by all MSs in cell) and the dedicated (used by individual MSs) channel categories apply to physical, transport, and logical channels. The shared (shared by some MSs in a cell) channel category applies to physical and transport channels only. Shared channels typically require some form of random access procedure, and use RNTI (Section 13.4.6) to identify MSs.

UTRAN Signaling Messages. UTRAN signaling messages (in the control plane) can be of two types:

1. Messages that are used to build and manage the radio connection between the MS and the BS (the Access Stratum—Section 13.4.1). They are generated by the peer RRC functional entities in the MS and in the BS.
2. Messages that are used for call control and mobility management (the Non-Access Stratum, see Section 13.4.1), which are generated by higher layer entities in the mobile station and the CN [27]. These messages flow transparently through the BS, going directly between the mobile station and the mobile switching center.

This chapter covers signaling in the Access Stratum.

13.4.3 Physical Layer

The physical layer provides data transport between the MS and BS and manages frequencies, timing, modulation, demodulation, coding, spreading, despreading, and power control [36–40].

UTRAN uses a frequency band of 5 MHz and a chip rate of 3.84 Mc/s for both the uplink and the downlink channel.

The physical layer interfaces with the MAC sublayer via transport channels and with the peer physical layer (at the far end of radio link) via physical channels. Physical channels are listed in Table 13.4-1; transport channels and the physical channels they map into are shown in Table 13.4-2. Each transport channel maps into a physical channel, and more than one transport channel may map into the same physical channel. The physical channels that do not map into transport channels perform layer 1 signaling functions only.

Physical channels are separated by *spreading codes* and *scrambling codes.* Each channel's baseband signal is first spread with a spreading (*channelization*) code and then multiplied by the scrambling code. A high-level conceptual view of how physical channels are realized is shown in Fig. 13.4-3. Figure 13.4-3(*a*) shows a downlink channel (except SCH) and Fig. 13.4-3(*b*) shows a group of uplink channels.

Scrambling Codes. These codes are PN sequences used in the uplink channel to separate transmissions from individual MSs, and in the downlink channel to separate cells or cell sectors.

The uplink channel uses short or long scrambling codes, based on channel type: some channels use only long codes, while some may use either long or short codes. The choice, for the channels that allow it, is an implementation decision. Long codes are based on *Gold code* sequences with a length of $2^{25}-1$ chips, of which only a section (38,400 chips) is used. Short codes are based on *S*(*2*) *codes* with a length of 256 chips. Each mobile station uses several scrambling codes, out of a pool of 2^{24} ($\sim$16.8 M).

The downlink channel uses long codes, based on Gold code sequences with a length of $2^{18}-1$ chips, of which only a section (38,400 chips) is used. Codes are picked from a pool of 8192 codes, out of a theoretical total of 262,143. The pool of codes is divided into 512 sets made of one primary and fifteen secondary scrambling codes, and each cell uses one set. The number of primary codes is kept small to speed up cell search, and they are arranged into 64 *code groups*, each containing eight primary codes.

Spreading Codes. These codes are used to separate individual code channels in both the downlink and uplink directions, within the broader channels created by scrambling codes. To provide flexible data rates, UTRAN uses *orthogonal variable spreading factor* (OVSF) codes, which are orthogonal Walsh codes of variable length. Walsh codes in an equal-length set are orthogonal, but Walsh codes of different lengths are not necessarily so. OVSF is a tree-based algorithm that generates sets

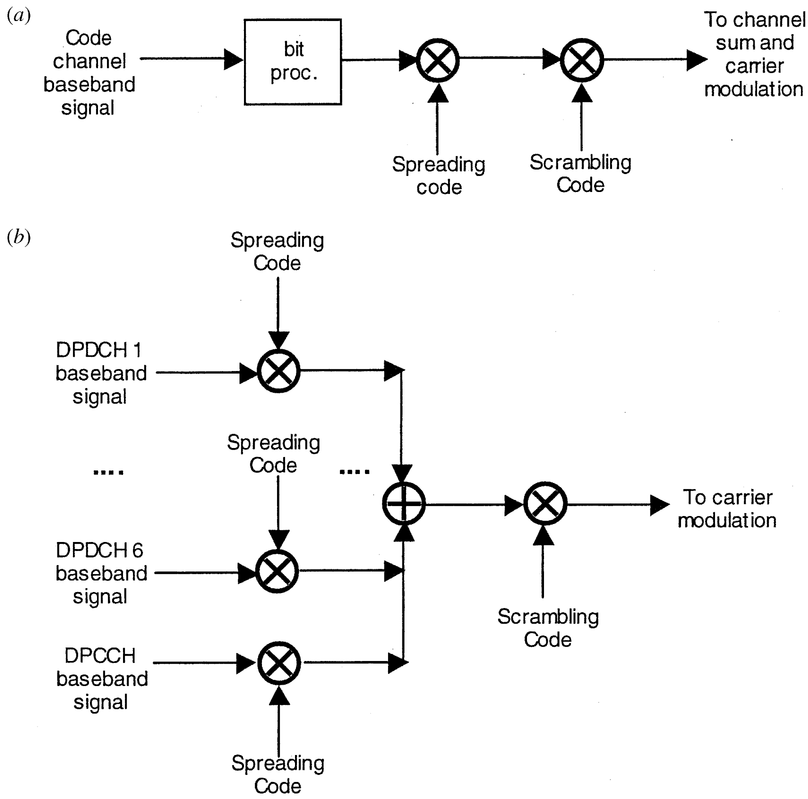


Figure 13.4-3. Conceptual view of bit processing for UTRAN physical channels.

of variable-length Walsh codes that are orthogonal and thus can be used for channelization.

Frames and Slots. Physical channel transmissions are structured as a sequence of *frames*, and each frame is subdivided into a sequence of *slots*. Some channels also group slots into three-slot *subframes*. Frames, subframes, and slots are the basic timing intervals for power control, data rate adjustment, and so on. Most channel frames are 10 ms long and have 15 slots, corresponding to 38,400 chips per frame and 2560 chips per slot (Fig. 13.4-4). Examples of slot formats in the uplink and the downlink channel are shown in Fig. 13.4-5. The TPC and FBI fields are used for power control and the TFCI field carries format information such as the spreading factor.

Figure 13.4-5 shows how the downlink DPCH time-multiplexes user (DPDCH) and control (DPCCH) data in the same code channel, but the equivalent data streams are carried in two separate code channels in the uplink. The time-multiplexed configuration is more efficient but can produce EMC interference

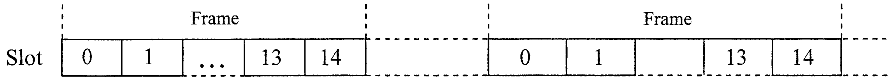


Figure 13.4-4. UTRAN frames.

during silent pauses, when no speech-carrying bits are transmitted. During such pauses, control-data transmission (e.g., power control) does not stop and generates pulses in a frequency range that can interfere with electrical equipment. The problem is serious only in the uplink direction because MSs may be near susceptible appliances such as hearing aids.

Physical Channels. ("Mapping" refers to Transport Channels.)

AICH: Acquisition Indicator Channel. This downlink common channel is used by the BS to acknowledge receipt of an access request on the PRACH.

DPCH: Dedicated Physical Channel. This downlink dedicated channel maps into the downlink DCH. The DPCH frame contains the downlink DPCCH and DPDCH subchannels, time multiplexed. The equivalent channels in the uplink, DPCCH and DPDCH, each have their own primary channel.

E-AGCH: E-DCH Absolute Grant Channel. This downlink common channel carries control information (*absolute grants*) for the E-DCH.

E-HICH: E-DCH Hybrid ARQ Indicator Channel. This downlink dedicated channel carries control information (*ARQ acknowledgments*) for the E-DCH.

E-RGCH: E-DCH Relative Grant Channel. This downlink dedicated channel carries control information (*relative grants*) for the E-DCH.

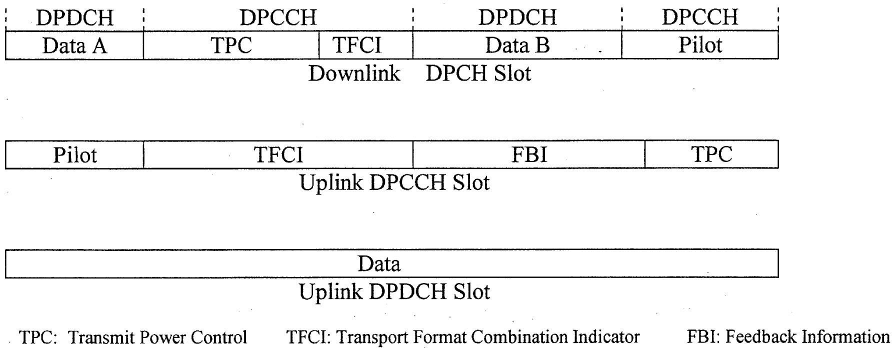


Figure 13.4-5. Examples of slot formats.

F-DPCH: Fractional Dedicated Physical Channel. This downlink dedicated channel carries control information. It has a function equivalent to the downlink DPCCH.

HS-PDSCH: High-Speed Physical Downlink Shared Channel. This downlink shared channel maps into the HS-DSCH. One HS-DSCH can map into multiple HS-PDSCH. Multiplexing of data from different MSs into one HS-PDSCH is possible. It is always paired with a DPCH and works in concert with HS-SCCH and HS-DPCCH.

HS-SCCH: High-Speed Shared Control Channel. This downlink common channel is used for control of HS-PDSCH. It is used only by the physical layer.

MICH: MBMS Indicator Channel. This downlink common channel carries MBMS (multimedia broadcast multicast service [41]) notification indicators.

P-CCPCH: Primary Common Control Physical Channel. This downlink common channel maps into the BCH. It is monitored by all UEs in a cell and is time-multiplexed with SCH.

P-CPICH/S-CPICH: Primary/Secondary Common Pilot Channel. This downlink common channel is used by the MS as a reference for signal strength measurements needed for cell acquisition and handover. It is divided into two subchannels: *primary*, covering the entire cell, and *secondary*, used with narrow beam antennas. A cell has one primary channel and optionally several secondary channels. It is used only by the physical layer.

PICH: Paging Indicator Channel. This downlink common channel is used by the BS to alert a group of MSs that a paging message for the paging group to which the MSs belong is expected on the PCH. It works in concert with S-CCPCH. It allows MSs to save power by ignoring paging for other groups. It is used only by the physical layer.

S-CCPCH: Secondary Common Control Physical Channel. This downlink common channel maps into FACH and PCH. FACH and PCH can share the same S-CCPCH or use separate ones.

SCH: Synchronization Channel. This downlink common channel is used by the MS to synchronize with the BS during cell search. It is divided into primary and secondary subchannels by means of different codes. It is used only by the physical layer.

DPCCH: Dedicated Physical Control Channel. This uplink dedicated channel carries layer 1 control data for the DPDCH. One DPCCH channel can be assigned per MS. It is used only by the physical layer.

DPDCH: Dedicated Physical Data Channel. This uplink dedicated channel maps into the uplink DCH. It supports variable data rates. Several DPDCH channels can be assigned to a MS, under the control of one DPCCH.

E-DPCCH: E-DCH Dedicated Physical Control Channel. This uplink dedicated channel is used to control the E-DPDCH. One E-DPCCH can be assigned per MS.

E-DPDCH: E-DCH Dedicated Physical Data Channel. This uplink dedicated channel maps into the E-DCH. Several E-DPDCH channels can be assigned to a MS.

HS-DPCCH: High Speed Dedicated Physical Control Channel. This uplink dedicated channel is used by the MS to send feedback information about HS-PDSCH and HS-SCCH, such as ARQ and channel quality data, to the BS. One HS-DPCCH channel can be assigned to a MS. It is used only by the physical layer.

PRACH: Physical Random Access Channel. This uplink common channel maps into the RACH. It uses a random access protocol.

Transport Channels

BCH: Broadcast Channel. This downlink common channel is used to broadcast system information, such as access codes, to all the MSs in a cell.

FACH: Forward Access Channel. This is a downlink common channel whose primary function is to send control information to MSs in response to a RACH request. MS identification is provided during communication. It may be used also for small user-data packet transmissions.

HS-DSCH: High-Speed Downlink Shared Channel. This downlink shared channel can be used as an extension of DCH and is shared dynamically by several MSs.

PCH: Paging Channel. This downlink common channel is used by the BS for paging and to broadcast control information.

DCH: Dedicated Channel. This downlink and uplink dedicated channel is used for voice, signaling, and extended data sessions. It supports variable data rates. Multiple DCHs can be set up for the same MS to support services such as voice and video.

E-DCH: Enhanced Dedicated Channel. This uplink dedicated channel can be used to extend DCH transmission.

RACH: Random Access Channel. This is an uplink common channel whose primary function is to transmit requests for system access to the BS. It may also be used for small user-data packet transmissions.

Transport Blocks. Transport blocks are the units of data exchanged between the MAC sublayer and the physical layer, which the physical layer maps into frames. The physical layer adds a CRC field when transmitting and removes it when receiving.

13.4.4 MAC Sublayer

The *medium access control* (MAC) sublayer manages communication between the physical layer and the RLC sublayer [42]; it is connected to the physical layer by transport channels and to the RLC by logical channels (Table 13.4-3). The MAC sublayer maps transport channels to logical channels and handles multiplexing/demultiplexing of upper layer PDUs into/from physical layer transport blocks.

Not all MAC PDUs exchanged with the physical layer need headers, depending on the type of logical channel(s) carried in the PDUs. A MAC header is needed when the same logical channel can map into different transport channels (see Fig. 13.4-6 for the most common format). In that case, the carried logical channels are identified in the *target channel type field* (TCTF) and the C/T field of the MAC header.

Logical Channels

BCCH: Broadcast Control Channel. This downlink common channel is used for broadcasting system information to a cell.

CCCH: Common Control Channel. This downlink and uplink common channel is used to send signaling information between MSs and the BS when the MS is not connected to the network at the RRC sublayer (Section 13.4.6).

CTCH: Common Traffic Channel. This downlink common channel is used for sending user data to all or a specific set of MSs.

DCCH: Dedicated Control Channel. This downlink and uplink dedicated channel is used to send signaling information between the BS and an individual MS, when the MS is connected at the RRC sublayer (Section 13.4.6).

Header				RLC PDU
TCTF	IDT	ID	C/T	Data

TCTF: Target Channel Type Field IDT: Mobile Station ID type ID: Mobile Station ID or MBMS ID

Figure 13.4-6. MAC PDU format.

DTCH: Dedicated Traffic Channel. This downlink and uplink dedicated channel carries user traffic for an individual MS.

MCCH: MBMS Point-to-Multipoint Control Channel. This downlink shared channel carries control information for MSs set up to receive MBMS transmissions [41].

MSCH: MBMS Point-to-Multipoint Scheduling Channel. This downlink shared channel carries scheduling information for one or more MTCHs (see below).

MTCH: MBMS Point-to-Multipoint Traffic Channel. This downlink common channel carries user data for MBMS. It works in concert with MCCH and MSCH.

PCCH: Paging Control Channel. This downlink common channel carries paging information.

13.4.5 Radio Link Control (RLC) Sublayer

The RLC sublayer interfaces at its lower boundary with the MAC sublayer and, at its upper boundary, with layer 3 or upper sublayers of layer 2 [43,44]. The interfaces with the upper layers/sublayers are different in the control and the user planes (Fig. 13.4.2).

The major functions of the RLC sublayer are *segmentation and reassembly* and *QoS*.

Segmentation and reassembly breaks up longer upper layer PDUs, so they can fit into the smaller MAC PDUs, and reassembles them at reception.

Different levels of QoS are provided by RLC via three types of connection services to the RRC sublayer:

1. Acknowledged data transfer
2. Unacknowledged data transfer
3. Transparent data transfer

The first two types are equivalent to the services provided by cdma2000's ARQ sublayer (Section 13.3.5), while the third is specific to UTRAN.

Acknowledged data transfer handles segmentation and reassembly of higher-layer PDUs with guarantee of delivery; unacknowledged and transparent data transfer handle segmentation and reassembly without guarantee of delivery. Unlike the other two modes, transparent data transfer does not add a header and is used only when the higher-layer PDU can be reconstructed without header information (e.g., fixed-length PDU).

An example of RLC PDU format is shown in Fig. 13.4-7.

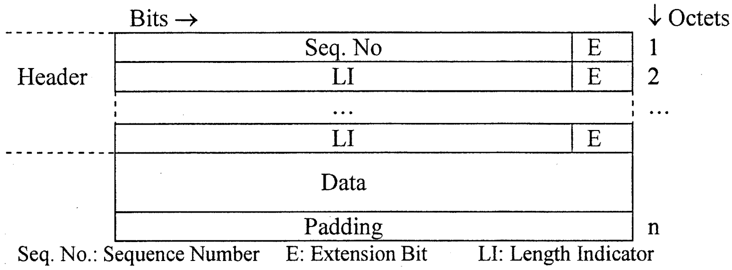


Figure 13.4-7. RLC PDU format for unacknowledged mode.

13.4.6 Radio Resource Control (RRC) Sublayer

The RRC sublayer is present only in the control plane, which is another way of saying that it only handles signaling traffic. User traffic does not go through the RLC sublayer (Fig. 13.4-2) [45]. The RRC interfaces with the RLC on the lower boundary and with higher-layer signaling on the other boundary. The main functions of RRC are:

- Management of signaling radio bearers
- Management of connections between RRC entities in MS and BS
- Control of QoS

RRC connections are bidirectional and a MS may have only one RRC connection established at a time.

The RRC sublayer provides local signaling connectivity between peer entities in the MS and the BS and manages the radio access portion of the network. Once established, RRC connections allow higher-layer signaling functions to exchange messages for call management and mobility management, functions that connect a MS directly with the core network (Fig. 13.4-1). Call management and mobility management messages are transferred transparently between the MS and CN, encapsulated into RRC *Direct Transfer* messages.

RRC supports two modes of operation:

1. Idle mode
2. Connected mode

In idle mode the MS is activated but not involved in a user-to-user connection. The MS is identified by global (non-access-network) identifiers such as IMSI or TMSI (Section 12.7.4).

In connected mode the MS is involved in a user-to-user connection. The MS is identified by a local identifier, called the *UMTS radio network temporary identifier* (U-RNTI or simply RNTI). The RNTI does not have any significance outside the radio access network.

TABLE 13.4-4 RRC Messages

Message	Logical Channel	Direction
Inter RAT Handover Information	N/A	MS → BS
Measurement Report	DCCH	MS → BS
Paging Type 1	PCCH	BS → MS
Paging Type 2	DCCH	BS → MS
Radio Bearer Release	DCCH	BS → MS
Radio Bearer Release Complete	DCCH	MS → BS
Radio Bearer Setup	DCCH	BS → MS
Radio Bearer Setup Complete	DCCH	MS → BS
RRC Connection Release	CCCH DCCH	BS → MS
RRC Connection Release Complete	DCCH	MS → BS
RRC Connection Request	CCCH	MS → BS
Signalling Connection Release	DCCH	BS → MS
Signalling Connection Release Indication	DCCH	MS → BS
Uplink Direct Transfer	DCCH	MS → BS

Note: RAT, radio access technology; MS, mobile station; BS, base station.

RRC signaling messages are listed in Table 13.4-4. UTRAN standards specify message formats in the ASN.1 notation (Section 16.2) [46], which makes it impractical to show their bit maps in concise form. The reader interested in the detailed formats is referred to [45].

As an example of message contents, Table 13.4-5 shows the most significant information elements in the uplink Direct Transfer message. The CN Domain Identity IE has two values: Circuit-Switched Domain and Packet-Switched Domain. The NAS message IE contains the message to/from the core network (the Non-Access Stratum).

13.4.7 UTRAN Call Sequences

Call sequences are similar to the ones described for IS-95 in Section 13.2.5, with the exception of the initialization procedure, which is markedly different and is described below

An example of a message sequence for the setup and release of a RRC connection is shown in Fig. 13.4-8.

TABLE 13.4-5 Information Elements in the Uplink Direct Transfer Message

IE
CN domain identity
Message authentication code
Message type
NAS message
RRC message sequence number
RRC transaction identifier

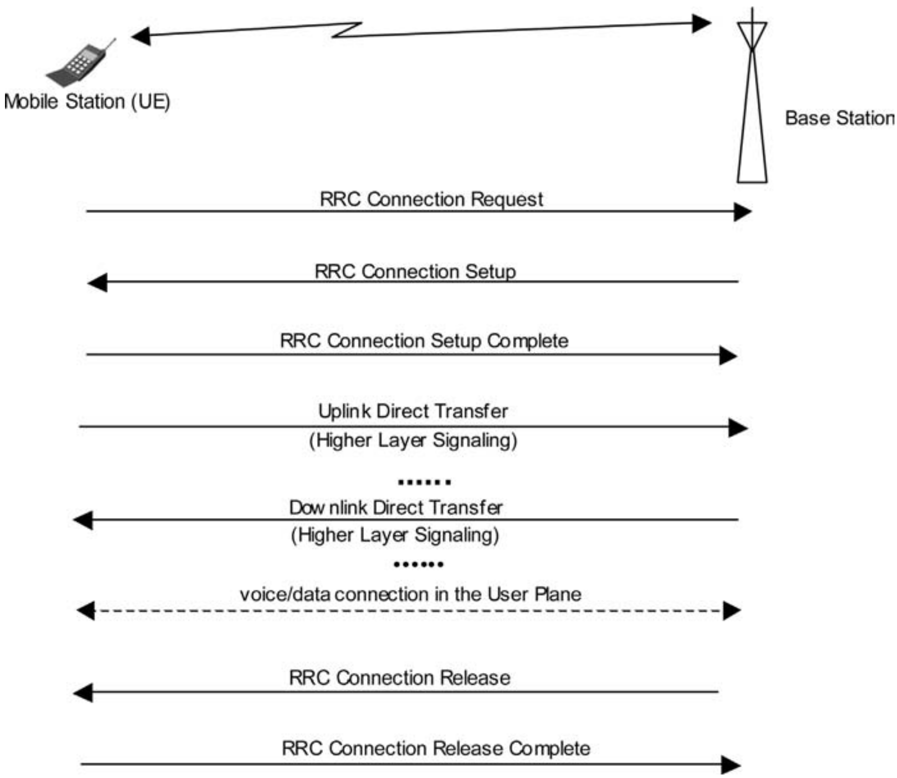


Figure 13.4-8. Example of RRC message sequence.

MS Initialization. UTRAN BSs are not synchronized, so they cannot use the phase offsets of a PN code (scrambling code) for BS separation. UTRAN systems must use different codes to identify different BSs.

The pilot and the synchronization channels (SCH, CPICH) are used for cell acquisition, like in IS-95 and cdma2000, but in a different order and following a three-step process.

The first step is the acquisition of slot synchronization, done by detecting the *primary synchronization code*, a 256-chip code (the same for all BSs in a system) transmitted continuously at the beginning of every slot in the primary synchronization channel (Fig. 13.4-9).

The second step is the acquisition of frame synchronization, done by detecting the *secondary synchronization codes*. From a pool of sixteen 256-chip codes, 64 code groups are obtained, each containing 15 of the 16 codes sequenced in a particular order. Each of the 64 secondary synchronization code groups maps into (identifies) a primary scrambling code group. A given secondary synchronization code group is repeated in every frame, with each code in the group transmitted sequentially at the beginning of each slot of the secondary synchronization channel

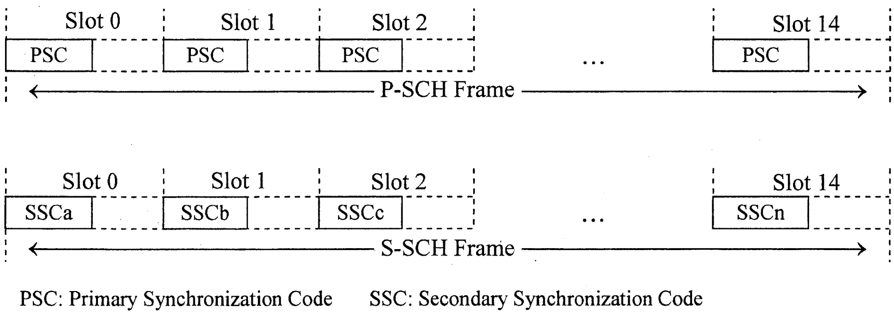


Figure 13.4-9. Primary and secondary synchronization channels.

(Fig. 13.4-9). The BS searches among all the code groups using a correlator circuit, until it finds a match. Acquisition of the secondary synchronization codes provides both frame synchronization (delineation of frame boundaries) and the identity of the primary scrambling code group.

The third step is the acquisition of the primary scrambling code from the pilot channel. The pilot channel transmits a continuous bit pattern (typically all zeros) spread with a known spreading code and scrambled with the primary scrambling code. The MS searches through the codes in the primary scrambling code group acquired in the second step, until it finds a match. The primary scrambling code allows the MS to decode the CCPCH (which also has a known spreading code) and to receive system information.

13.5 ACRONYMS

2G	Second generation
2.5G	Second generation and a half
3G	Third generation
3GPP	Third Generation Partnership Project
3GPP2	Third Generation Partnership Project 2
AMPS	Advanced mobile phone service
ANSI	American National Standards Institute
ARQ	Automatic repeat request
ASN.1	Abstract syntax notation one
AS	Access stratum
B/MS	Broadcast/multicast services
BMAC	Broadcast/multicast medium access control
BMC	Broadcast/multicast control
BS	Base station
BSC	Base station controller
BTS	Base transceiver station/system
CCMS	Common channel multiplex sublayer

CCSA	China Communications Standards Association
CDMA	Code-division multiple access
CN	Core network
CRC	Cyclic redundancy check
dB	Decibel
DS	Direct sequence, direct spread
EIA	Electronic Industry Alliance
ESN	Electronic serial number
ETSI	European Telecommunications Standards Institute
FCH	Fundamental Channel
FDD	Frequency-division duplex
FER	Frame error rate
FH	Frequency hopping
F-PDHCF	Forward packet data channel control function
GPS	Global positioning system
GSM	Global System for Mobile Communications
ID	Identity
IE	Information element
IMSI	International mobile subscriber identity
IMT-2000	International Mobile Telecommunications 2000
IP	Internet Protocol
IS	Interim standard
ITU	International Telecommunication Union
ITU-T	Telecommunications Standardization Sector of ITU
L1	Layer 1
L2	Layer 2
L3	Layer 3
LAC	Link access control
LPM	Logical-to-physical mapping
LSR	Linear-feedback shift register
MAC	Medium access control
MBMS	Multimedia broadcast multicast service
MC	Multicarrier
MCC	Mobile country code
MCSB	Message control and status block
MDN	Mobile directory number
MEID	Mobile equipment identifier
MIN	Mobile identification number
MM	Mobility management
MNC	Mobile network code
MS	Mobile station
MSC	Mobile switching center
MSIN	Mobile station identification number
NAS	Non-access stratum

NID	Network ID
ORDQ	Order qualification code
OSI	Open Systems Interconnection
PCS	Personal communications system
PDCP	Packet Data Convergence Protocol
PDHCF	Packet data channel control function
PDSN	Packet data service node
PDU	Protocol data unit
PN	Pseudorandom number
PUF	Power-up function
QoS	Quality of service
QPSK	Quadrature phase-shift keying
RAN	Radio access network
RAT	Radio access technology
RB	Radio bearer
RC	Radio configuration
RF	Radiofrequency
RLC	Radio link control
RNC	Radio network controller
RNS	Radio network subsystem
RNTI	Radio access network temporary identifier
R-PDHCF	Reverse packet data channel control function
RRC	Radio resource control
SAP	Service access point
SAR	Segmentation and reassembly
SDU	Service data unit
SID	System ID
SINR	Signal-to-interference ratio
SIR	Signal-to-interference ratio
SNR	Signal-to-noise ratio
SR	Spreading rate
SSD	Shared secret data
TDD	Time-division duplex
TDMA	Time-division multiple access
TIA	Telecommunications Industry Association
TMSI	Temporary mobile station identity
UE	User equipment
UIM	User identity module
ULS	Upper-layer signaling
USIM	User services identity module
UMTS	Universal Mobile Telecommunication System
U-RNTI	UMTS radio network temporary identifier
UTRAN	UMTS terrestrial radio access network
VoIP	Voice over IP
WCDMA	Wideband CDMA

13.6 REFERENCES

1. L. Harte, M. Hoenig, D. McLaughlin, and R. K. Kta, *CDMA IS-95 for Cellular and PCS Technology, Economics and Services*, McGraw-Hill, New York, 1999.
2. V. K. Garg, *IS-95 CDMA and cdma2000 Cellular/PCS Systems Implementation*, Prentice Hall PTR, Upper Saddle River, NJ, 2000.
3. S. Lee, *Spread Spectrum CDMA IS-95 and IS-2000 for RFC Communications*, McGraw-Hill, New York, 2002.
4. Alex Brand and Hamid Aghvami, *Multiple Access Protocols for Mobile Communications GPRS, UMTS and Beyond*, John Wiley & Sons, Chichester, England, 2002.
5. J. Viterbi, *CDMA: Principles of Spread Spectrum Communication*, Addison Wesley, Reading, MA, 1995.
6. D. Gerakoulis and E. Geraniotis, *CDMA: Access and Switching for Terrestrial and Satellite Networks*, John Wiley & Sons, Chichester, England, 2001.
7. *Mobile Station–Base Station Compatibility Standard for Wideband Spread Spectrum Cellular Systems*, ANSI/TIA/EIA-95-B, Telecommunications Industry Association, Arlington, VA, Feb. 1999.
8. V. Vanghi, A. Damnjanovic, and B. Vojcic, *The CDMA 2000 System for Mobile Communications*, Prentice Hall PTR, Upper Saddle River, NJ, 2004.
9. C. Smith and D. Collins, *3G Wireless Networks*, McGraw-Hill, New York, 2002.
10. *Capabilities Requirements Mapping for cdma2000 Standards*, C.R1000-0 Version 1.0, 3GPP2, Arlington, VA, Dec. 1999.
11. *3GPP2 System Capability Guide Release B*, S.R0003-A Version 1.0, 3GPP2, Arlington, VA, 14 June 2001.
12. *Introduction to cdma2000® Spread Spectrum Systems*, TIA-2000.1-D, Telecommunications Industry Association, Arlington, VA, Mar. 2004.
13. *Introduction to cdma2000 Spread Spectrum Systems Revision D*, C.S0001-D Version 1.0, 3GPP2, Arlington, VA, Feb. 2004.
14. *Network Reference Model for cdma2000 Spread Spectrum Systems Revision: B*, S.R0005-B Version 1.0, 3GPP2, Arlington, VA, 16 Apr. 2001.
15. *cdma2000 High Rate Packet Data Air Interface Specification*, C.S0024-A Version 1.0, 3GPP2, Arlington, VA, Mar. 2004.
16. *Wireless IP Architecture Based on IETF Protocols*, P.R0001 1 Version 1.0.0 2, 3GPP2, Arlington, VA, 14 July, 2000.
17. *Physical Layer for cdma2000® Spread Spectrum Systems*, TIA-2000.2-D, Telecommunications Industry Association, Arlington, VA, Mar. 2004.
18. *Physical Layer Standard for cdma2000 Spread Spectrum Systems Revision D*, C.S0002-D Version 1.0, 3GPP2, Arlington, VA, 13 Feb. 2004.
19. *Medium Access Control (MAC) Standard for cdma2000® Spread Spectrum Systems*, TIA-2000.3-D, Telecommunications Industry Association, Arlington, VA, Mar. 2004.
20. *Medium Access Control (MAC) Standard for cdma2000 Spread Spectrum Systems Release D*, C.S0003-D Version 1.0, 3GPP2, 13 Feb. 2004.
21. *Signaling Link Access Control (LAC) Standard for cdma2000® Spread Spectrum Systems*, TIA-2000.4-D, Telecommunications Industry Association, Arlington, VA, Mar. 2004.

22. *Signaling Link Access Control (LAC) Standard for cdma2000 Spread Spectrum Systems Release D*, C.S0004-D Version 1.0, 3GPP2, Arlington, VA, 13 Feb. 2004.
23. *Upper Layer (Layer 3) Signaling Standard for cdma2000® Spread Spectrum Systems*, TIA-2000.5-D, Telecommunications Industry Association, Arlington, VA, Mar. 2004.
24. *Upper Layer (Layer 3) Signaling Standard for cdma2000 Spread Spectrum Systems Release D*, C.S0005-D Version 1.0, 3GPP2, Arlington, VA, 13 Feb. 2004.
25. *Data Service Options for Wideband Spread Spectrum Systems*, TIA/EIA/IS-707-A/A1/A2/A3, Telecommunications Industry Association, Arlington, VA, 1999–2003.
26. *Administration of Parameter Value Assignments for cdma2000 Spread Spectrum Standards Release D*, C.R1001-D, 3GPP2, Arlington, VA, 18 Apr. 2003.
27. S. Low and R. Schneider, *CDMA Internetworking: Deploying the Open A-Interface*, Prentice Hall, Upper Saddle River, NJ, 2000.
28. T. Ojanpera and R. Prasad, *WCDMA: Towards IP Mobility and Mobile Internet*, Artech House, Boston, 2001.
29. B. Walke, P. Seidenberg, and M. P. Althoff, *UMTS The Fundamentals*, John Wiley & Sons, Chichester, England, 2003.
30. Harry Holma and Antti Toskala, *WCDMA for UMTS Radio Access for Third Generation Mobile Communications*, John Wiley & Sons, Chichester, England, 2000.
31. *Universal Mobile Telecommunications System (UMTS); UTRAN Overall Description (Release 6)*, TS 25.401 V6.6.0, 3GPP, Sophia Antipolis, France, June 2005.
32. *Universal Mobile Telecommunications System (UMTS); Numbering, Addressing and Identification (Release 6)*, TS 23.003 V6.7.0, 3GPP, Sophia Antipolis, France, June 2005.
33. *Universal Mobile Telecommunications System (UMTS);* Technical Specifications and Technical Reports for a UTRAN-Based 3GPP System, TS 21.101 V.6.2.0, 3GPP, Sophia Antipolis, France, June 2005.
34. *Universal Mobile Telecommunications System (UMTS); Radio Interface Protocol Architecture (Release 6)*, TS 25.301 V6.3.0, 3GPP, Sophia Antipolis, France, June 2005.
35. *3rd Generation Partnership Project; Technical Specification Group Services and System Aspects UMTS Access Stratum; Services and Functions (Release 5)*, TS 23.110 V6.0.0, 3GPP, Sophia Antipolis, France, Dec. 2004.
36. *Universal Mobile Telecommunications System (UMTS); Physical Layer—General Description (Release 6)*, TS 25.201 V6.2.0, 3GPP, Sophia Antipolis, France, June 2005.
37. *Universal Mobile Telecommunications System (UMTS); Physical Channels and Mapping of Transport Channels onto Physical Channels (FDD) (Release 6)*, TS 25.211 V6.5.0, 3GPP, Sophia Antipolis, France, June 2005.
38. *Universal Mobile Telecommunications System (UMTS); Multiplexing and Channel Coding (FDD) (Release 6)*, TS 25.212 V6.5.0, 3GPP, Sophia Antipolis, France, June 2005.
39. *Universal Mobile Telecommunications System (UMTS); Spreading and Modulation (FDD) (Release 6)*, TS 25.213 V6.3.0, 3GPP, Sophia Antipolis, France, June 2005.
40. *Universal Mobile Telecommunications System (UMTS); Services Provided by the Physical Layer (Release 6)*, TS 25.302 V6.4.0, 3GPP, Sophia Antipolis, France, June 2005.
41. *Introduction of the Multimedia Broadcast Multicast Service (MBMS) in the Radio Access Network (RAN) (Stage-2)*, TS 25.346, 3GPP, Sophia Antipolis, France, June 2005.

42. *Universal Mobile Telecommunications System (UMTS); Medium Access Control (MAC) Protocol Specification (Release 6)*, TS 25.321 V6.5.0, 3GPP, Sophia Antipolis, France, June 2005.
43. *Universal Mobile Telecommunications System (UMTS); Mobile Radio Interface Signalling Layer 3; General Aspects (Release 6)*, TS 24.007 V6.4.0, 3GPP, Sophia Antipolis, France, Mar. 2005.
44. *Universal Mobile Telecommunications System (UMTS); Radio Link Control (RLC) Protocol Specification (Release 6)*, TS 25.322 V6.4.0, 3GPP, Sophia Antipolis, France, June 2005.
45. *Universal Mobile Telecommunications System (UMTS); Radio Resource Control (RRC); Protocol Specification (Release 6)*, TS 25.331 V6.6.0, 3GPP, Sophia Antipolis, France, June 2005.
46. *Information Technology—Abstract Syntax Notation One (ASN.1): Specification of Basic Notation*, Rec. X.680, ITU-T, Geneva, 2002.

14

INTRODUCTION TO TRANSACTIONS

The TUP and ISUP protocols (Chapters 9 and 11) are SS7 protocols for trunk-related exchange actions (mainly call control). SS7 signaling takes place between the exchanges at the ends of a TUP or ISUP trunk. Starting with this chapter, the focus is on SS7 signaling for operations that are not trunk-related.

14.1 DEFINITIONS AND APPLICATIONS

14.1.1 Definitions

Remote Operation. A remote operation or, as most often called, an *operation* is a specifically defined action or set of actions that a user application (client) requests a provider application (server) to execute. Remote operations are not trunk-related and, as the name implies, the entity requesting them is in one node, and the entity executing them is in another node.

ITU-T has defined a protocol for *remote operations service* (ROS): the Remote Operations Service Element (ROSE) [1–5]. The ROSE protocol provides a generic definition of the parameters needed for operation-related communication between nodes. Operation requests and responses are based on five parameter fields:

- Operation code
- ARGUMENT {parameter 1, . . . , parameter m}
- RESULT {parameter 1, . . . , parameter n}
- ERRORS {Error 1, . . . , Error x}
- LINKED {Operation 1, . . . , Operation y}

Signaling in Telecommunication Networks, Second Edition, by John G. van Bosse and Fabrizio U. Devetak

Applications define operations by controlling which parameter fields are used and by assigning application-specific values to them.

Transaction. A transaction is a dialogue consisting of signaling messages between two signaling points (nodes) for the execution of one or more remote operations [6–8].

Transactions can involve two exchanges, an exchange and a network database, an exchange and a maintenance center, and so on. In this chapter we shall use the term "node" for those entities.

14.1.2 Applications

The applications that use transactions fall into several groups:

Intelligent network (IN) services such as "800" and "900" calls. For these calls an exchange needs to query a network database to obtain a routing number that corresponds to the received 800 or 900 number. Intelligent network services are discussed in Chapters 17 and 18.

Services for mobile telecommunications [8]. Transactions may be used to keep track of the present location of each active mobile station (MS). This information is used when routing a call to a MS. Services for mobile telecommunications are described in Chapter 19.

Centralized operation, administration, and maintenance (OAM). An OAM center uses transactions to verify and change the data stored in exchanges, to request a test of network equipment, and so on.

Bulk data transfer. This group of transactions handles, for example, the daily transfer of billing records from an exchange to a centralized revenue accounting center.

The transactions in the first two groups support the setup of calls and are known as "on-line" transactions. They have to be executed with minimal delays, since they add to the overall time to set up a call and usually require two short messages.

The transactions in the third and fourth groups are "off-line." They are less time-critical and usually require more and longer messages.

14.2 SS7 ARCHITECTURE FOR TRANSACTIONS

14.2.1 Application Entities and Application Service Elements

Nodes are hosts to applications and signaling point functions. In the ITU-T model for Signaling System No. 7, the part of an *application process* (AP) that handles the communication aspects of the application is called an *application entity* (AE). An AE contains one or more functions called *application service elements* (ASEs)

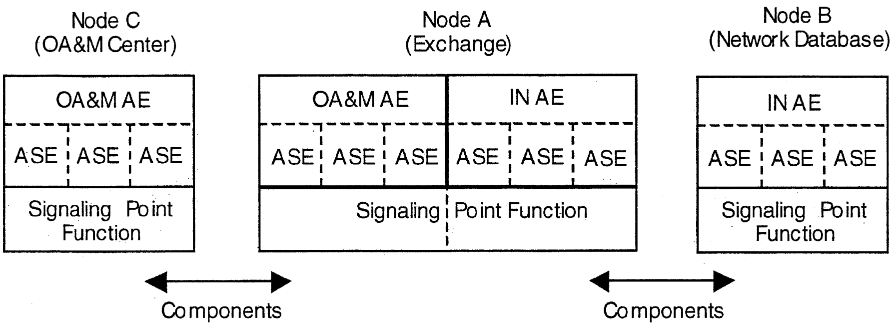


Figure 14.2-1. Application entities.

[9–11], which are typically a collection of operations defined in such a way that they can be used in multiple AEs. TCAP (Chapter 16), although not a group of operations, is also an ASE, common to AEs that use transactions. Transactions are handled by ASEs under the overall coordination of AEs.

In Fig. 14.2-1, node A is an exchange equipped with two AEs, to handle two types of transactions. The IN AE handles transaction messages for IN services, and the OAM AE handles transaction messages for operation, administration, and maintenance (OAM). The IN AE and OAM AE in node A communicate with their respective peers in an IN database (node B), and in an OAM center (node C). For the sake of the example, all the AEs in the figure have three ASEs, but the number of ASEs in an AE is an implementation choice.

The information elements (IEs) exchanged between ASEs in transaction messages for the purpose of executing operations are called *components*. Components carry the parameters of the operations, under the control of ASEs: ASEs decide which parameters are included in each type of component and what their values are.

14.2.2 Infrastructure for Transactions

Figure 14.2-2 shows the SS7 entities at a signaling point that are involved in the transfer of transaction messages.

The ASEs in an AE at an IN or OAM node are users of the Transaction Capabilities Application Part (TCAP) of Signaling System No. 7 and are called *TC-users*. TCAP is a user of the Signaling Connection Control Part (SCCP), which in turn uses the services of the Message Transfer Part (MTP).

In the ITU-T model of SS7, an ASE passes its outgoing components to the TCAP ASE in TC request primitives. TCAP places one or more components in a TCAP message and passes them to SCCP in an *N*-unitdata request primitive. SCCP then builds a SCCP message and passes it to MTP in a MTP-transfer request primitive. Finally, MTP forms a message signal unit (MSU), which is sent out on a signaling link.

Incoming components arrive in MSUs on a signaling link. MTP extracts the SSCP message and passes it to SCCP in a MTP-transfer indication primitive.

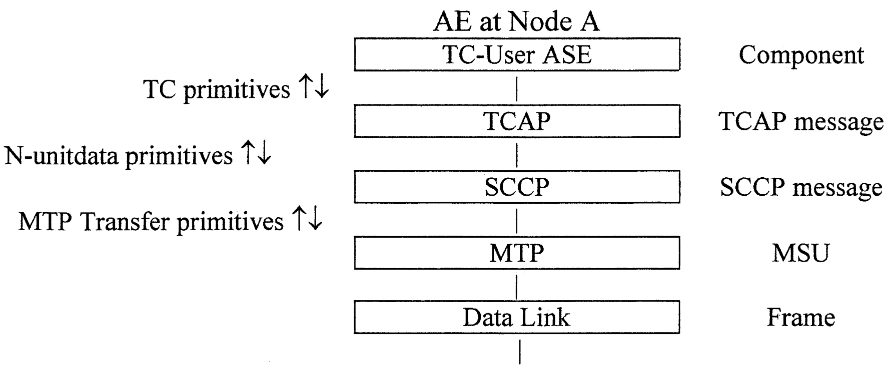


Figure 14.2-2. Architecture of transaction.

SCCP then extracts the TCAP package and passes it to TCAP in an *N*-unitdata indication primitive. Finally, TCAP extracts the component(s) and passes them to the appropriate AE, in TC indication primitives.

MTP, SCCP, and TCAP are described in Chapters 8, 15, and 16, respectively.

14.2.3 Application Independence of TCAP

In Fig. 14.2-1, most ASEs in the IN and OAM AEs are application-specific. The TCAP serves all ASEs at a node and is an application-independent protocol.

14.2.4 Identification of AEs and ASEs

An AE is identified by two parameters: the address of the node and the address of the service access point (SAP), that is, the address of the application within the node. In a telecommunication network using SS7 signaling, those two parameters are the *point code* (PC) of the signaling point (Section 7.2.1), and the *subsystem number* (SSN) of the SCCP user (Section 15.1). The combination PC + SSN is known as the "SCCP address" and is used by SCCP to deliver transaction messages to the destination AEs. Within an AE a transaction is directed to the appropriate ASE by the coordinating function of the AE, based on the invocation of a specific operation. Implementations have to ensure that a given operation appears only once in an AE.

14.3 ACRONYMS

AE	Application entity
ANSI	American National Standards Institute
AP	Application process
ASE	Application service element

CCIS	Common–channel interoffice signaling
CCITT	International Telegraph and Telephone Consultative Committee
IN	Intelligent network
ISDN	Integrated Services Digital Network
ISUP	ISDN User Part
ITU	International Telecommunication Union
ITU-T	Telecommunications Standardization Sector of ITU
MS	Mobile station
MSU	Message signal unit
MTP	Message Transfer Part
OAM	Operation, administration, and maintenance
OMASE	ASE for OAM applications
PC	Point code
ROS	Remote operations service
ROSE	Remote Operations Service Element
SAP	Service access point
SCCP	Signaling Connection Control Part
SSN	Subsystem number
SS7	Signaling System No. 7
TC	Transaction Capabilities
TCAP	Transaction Capabilities Application Part
TUP	Telephone User Part

14.4 REFERENCES

1. *Remote Operations: Model, Notation and Service Definition*, Rec. X.219, ITU-T, Geneva, 1988.
2. *Remote Operations: Protocol Specification*, Rec. X.229, ITU-T, Geneva, Nov. 1988.
3. *Remote Operations: Concept, Model and Notation*, Rec. X.880, ITU-T, Geneva, 1994.
4. *Remote Operations: OSI Realizations—Remote Operations Service Element (ROSE) Service Definition*, Rec. X.881, ITU-TM, Geneva, 1994.
5. *Remote Operations: OSI Realizations—Remote Operations Service Element (ROSE) Protocol Specification*, Rec. X.882, ITU-T, Geneva, 1994.
6. A. R. Modaressi and R. A. Skoog, Signaling System No. 7: A Tutorial, *IEEE Commun. Mag.*, **28**, No. 7, July 1990.
7. R. Mantefield, *Common Channel Signalling*, Peter Peregrinus Ltd., London, 1991.
8. T. W. Johnson, B. Law, and P. Anius, CCITTT Signalling System No. 7: Transaction Capabilities, *Br. Telecommun. Eng.*, **7**, Apr. 1988.
9. *Introduction to CCITT Signalling System No. 7*, Rec. Q.700, ITU-T, Geneva, 1993.
10. *Functional Description of Transaction Capabilities*, Rec. Q.771, ITU-T, Geneva, 1997.
11. *Guidelines for Using Transaction Capabilities*, Rec. Q.775, ITU-T, Geneva, 1997.

15

SIGNALING CONNECTION CONTROL PART

The Message Transfer Part (MTP) of SS7 (Chapter 8) was designed to transfer TUP (and later on, ISUP) messages between exchanges at the ends of a trunk. The Signaling Connection Control Part (SCCP), in combination with MTP, provides the transfer of messages that are not related to individual trunks, for example, transaction messages [1–3]. The combination of MTP and SCCP is known as the *network service part* (NSP) of SS7 and is the equivalent of layers 1, 2, and 3 of the OSI (open systems interconnection) protocols in data communication systems (Section 7.1.1).

SCCP has been documented by ITU-T, in a series of Recommendations [4–7]. The U.S. version has been specified by ANSI [8].

15.1 INTRODUCTION

Figure 15.1-1 shows SCCP and its relations with the other parts of SS7. TCAP and ISUP are SCCP users. In turn, the ASEs at a signaling point are users of TCAP and can be considered as "indirect" SCCP users. Each SCCP user at a signaling point has a *subsystem number* (SSN) that ranges 1 through 127. When SCCP receives an incoming message from MTP, it uses SSN to deliver the message to the appropriate SCCP user (in the case of ASEs, the message is passed to TCAP, which then delivers it). In this chapter, the term *subsystem* is used to denote a SCCP user.

15.1.1 Message Transfer Enhancements

SCCP enhances the message transfer capabilities of SS7 in two principal ways. In the first place, MTP uses the service indicator (SI) (Section 8.8.3). In an incoming message to deliver it to the appropriate MTP user (SCCP is one of them). SI has

Signaling in Telecommunication Networks, Second Edition, by John G. van Bosse and Fabrizio U. Devetak

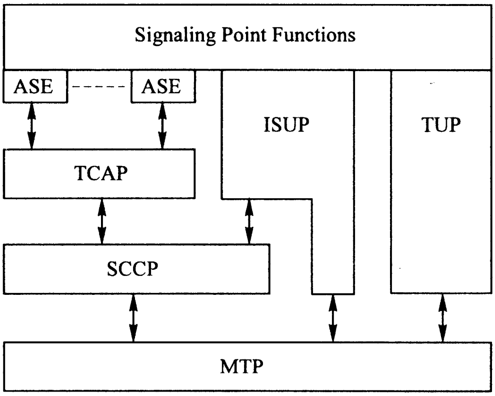


Figure 15.1-1. Position of SCCP in Signaling System No. 7. (From Rec. Q.700. Courtesy of ITU-T.)

a range 0 through 15, limiting the number of MTP users to 16. SCCP expands the addressing capability of SS7, allowing up to 127 subsystems at a signaling point.

In the second place, SCCP also allows a calling subsystem to address a called subsystem by a *global title* (GT). This is a functional address, in the form of a digit string, that cannot be used for message routing. The purpose of GTs is discussed in Section 15.3. SCCP includes provisions to translate a global title into an address that can be used to route a message to its destination subsystem.

15.1.2 SCCP Message Transfer Services

SCCP provides four service classes to its users:

- Class 0: Basic connectionless service
- Class 1: Connectionless service with sequence control
- Class 2: Basic connection-oriented service
- Class 3: Connection-oriented service with flow control

In connectionless services, the SCCPs at the signaling points of the two subsystems involved in a transaction are not aware of the transaction. In connection-oriented service, a (virtual) "signaling connection" between the SCCPs is established first, making both SCCPs aware of the transaction. At the end of the transaction, the signaling connection is released.

Messages for on-line transactions (Section 14.1) are usually transferred by connectionless SCCP service. Connection-oriented service is intended primarily for the transfer of messages in off-line transactions that involve a large amount of data. In addition, one group of on-line transactions in the Global System for Mobile Communications (GSM) uses connection-oriented SCCP—see Chapter 19.

15.1.3 Structure and Interfaces of SCCP

SCCP consists of the parts shown in Fig. 15.1-2. Connectionless control (SCLC) and connection-oriented control (SCOC) handle the message transfer in the corresponding services.

Routing control (SCRC) determines the destinations of outgoing messages and routes incoming messages to SCLC or SCOC, which then deliver them (directly or via TCAP) to the destination subsystems.

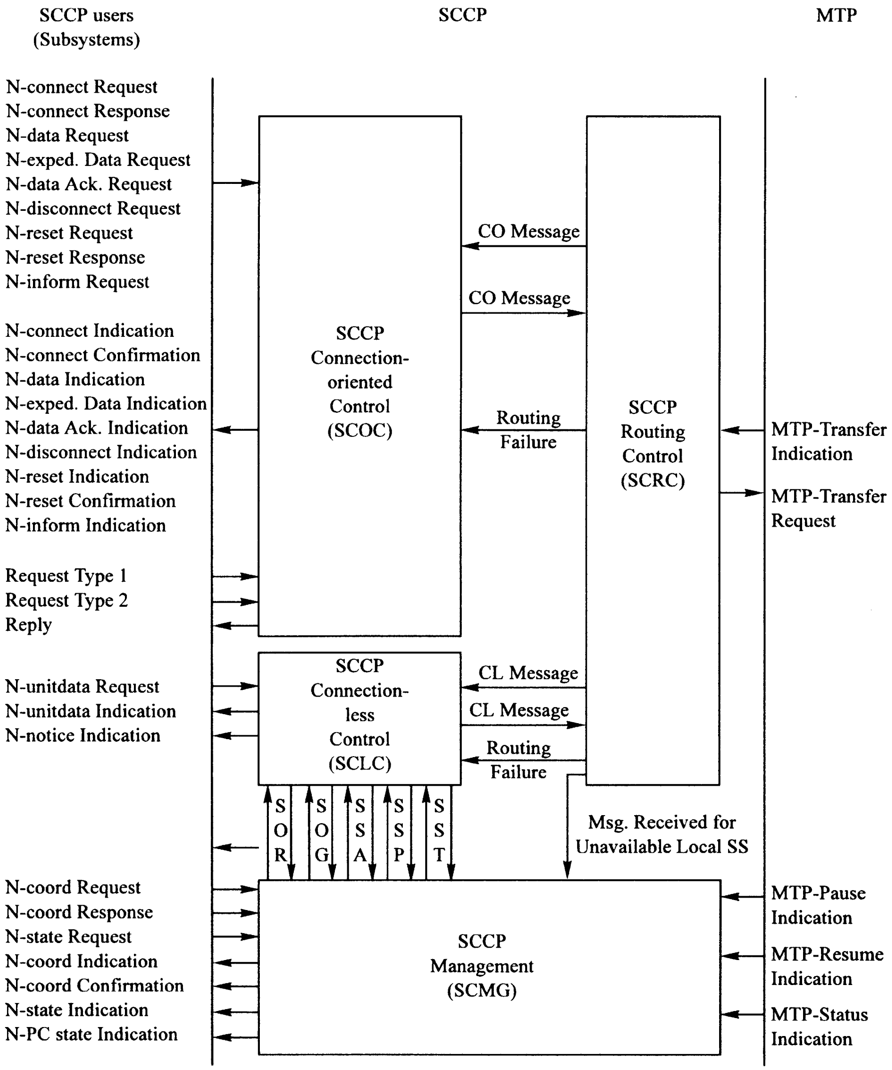


Figure 15.1-2. Structure and interfaces of SCCP. (From Rec. Q.714. Courtesy of ITU-T.)

SCCP management (SCMG) has functions comparable to those of MTP3 signaling network management (Section 8.8). It aims to maintain the message traffic of the SCCP users under congestion and failure conditions in the signaling network, signaling points, and subsystems.

The interfaces between SCCP and MTP are as described in Sections 8.8 and 8.9. MTP-transfer primitives are the interface between SCCP routing control and MTP for message transfer. MTP management passes MTP-pause, MTP-resume, and MTP-status indications to SCCP management.

N-Primitives form the interface between SCCP and its local subsystems (either directly or via TCAP). We distinguish three groups of primitives. The group to/from connection-oriented control, the group to/from connectionless control, and the group to/from SCCP management.

15.2 SCCP MESSAGES AND PARAMETERS

This section discusses a number of SCCP messages for the transfer of information between subsystems. SCCP messages are located in the user message (UM) fields of MTP3 messages (Fig. 8.8-1).

15.2.1 General Message Format

The format of SCCP messages is very similar to that of ISUP messages (see Fig. 11.2-2, without the CIC octets). A message can have mandatory (M) parameter, with fixed or variable lengths, and optional (O) parameters. As in ISUP messages (Section 11.2), the mandatory fixed-length parameters appear with their *value* (*contents*) field only. The mandatory variable-length parameters have a pointer and length and value fields. Each optional parameter has a name, length, and value field. The name field is also known as the *tag* field.

15.2.2 Message Types

The messages that will be discussed in this chapter are outlined below [5]. The first two apply to connectionless SCCP service; the others are used in connection-oriented service.

Unitdata Message (*UDT*). This is sent by an SCCP, to transfer subsystem data.

Unitdata Service Message (*UDTS*). This is sent to the SCCP that originated a UDT message, by an SCCP that cannot deliver a received UDT message to its destination.

Connection Request Message (*CR*). This is a request from a calling SCCP to a called SCCP, to setup a signaling connection.

Connection Confirm Message (*CC*). This is sent by called SCCP, indicating that it has set up the signaling connection.

TABLE 15.2-1 Message Type Codes of SCCP Messages

Acronym	Name	Message Type Code
CC	Connection Confirm	0000 0010
CR	Connection Request	0000 0001
CREF	Connection Refused	0000 0011
DT1	Data Form 1	0000 0110
DT2	Data Form 2	0000 0111
RLC	Release Complete	0000 0101
RLSD	Released	0000 0100
UDT	Unitdata	0000 1001
UDTS	Unitdata Service	0000 1010

Source: Rec. Q.713. Courtesy of ITU-T.

Connection Refused Message (*CREF*). This is sent by the called SCCP, indicating that it is unable to set up the signaling connection.

Data Form 1 (*DT1*). This is a message sent by a SCCP at either end of a signaling connection and contains subsystem data (used in class 2 operation).

Data Form 2 (*DT2*). This is a message sent by a SCCP at either end of a signaling connection. It contains subsystem data and acknowledges the receipt of messages (class 3 operation).

Released Message (*RLSD*). This is sent by SCCP at one end of the signaling connection and indicates that it wants to release the signaling connection.

Release Complete Message (*RLC*). This is sent in response to a RLSD message and indicates that the sending SCCP has released the connection.

The message type codes of these messages are listed in Table 15.2-1.

15.2.3 Message Parameters

This section describes the most important parameters in the above messages. At this point it is suggested merely to peruse the descriptions, and to refer back to them when reading the later sections of this chapter.

To allow the reader to locate a parameter description quickly, each parameter has a reference number (e.g., Par.1). In the sections that follow, a parameter is always identified by name and reference number.

Table 15.2-2 lists the parameters and the messages in which they appear.

The focus is on parameter contents, and only the "value" fields are shown in the figures. The name (tag) fields of parameters that can appear as optional parameters are coded as follows:

Par.1	Called Party Address:	0000 0010
Par.2	Calling Party Address:	0000 0011
Par.12	User Data:	0000 1111

TABLE 15.2-2 Parameters in SCCP Messages

		Message Acronyms								
Reference Number	Parameter Name	CC	CR	CREF	DT1	DT2	RLC	RLSD	UDT	UDTS
Par.1	Called party address (CDA)	O	M	O					M	M
Par.2	Calling party address (CGA)		O						M	M
Par.3	Destination local reference	M		M	M	M	M	M		
Par.4	Protocol class	M	M						M	
Par.5	Refusal cause		M							
Par.6	Release cause							M		
Par.7	Return cause									M
Par.8	Return option								M	
Par.9	Segmenting/reassembling				M					
Par.10	Sequencing/segmenting					M				
Par.11	Source local reference	M	M				M	M		
Par.12	SCCP user data (subsystem data)	O	O	O	M	M		O	M	M

		Service Class			
Acronym	Message	0(CL)	1(CL)	2(CO)	3(CO)
CC	Connection confirm			X	X
CR	Connection request			X	X
CREF	Connection refused			X	X
DT1	Data form 1			X	
DT2	Data form 2				X
RLC	Release complete			X	X
RLSD	Released			X	X
UDT	Unitdata	X	X		
UDTS	Unitdata service	X	X		

CL, Connectionless service; CO, connection-oriented service.

Source: Rec. Q.712. Courtesy of ITU-T.

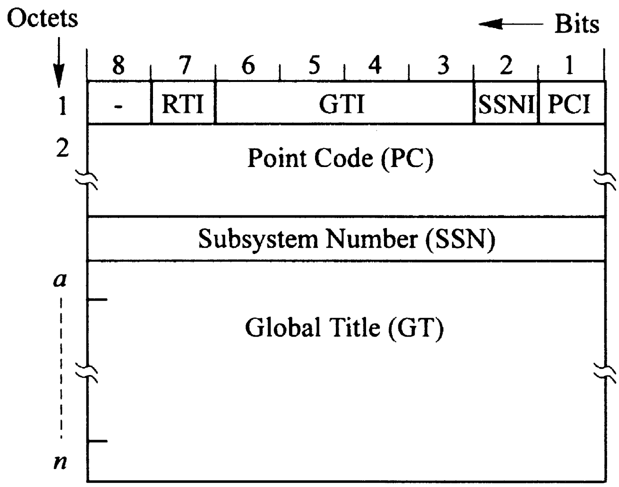


Figure 15.2-1. General formats of called party address (Par.1) and calling party address (Par.2). PCI, Point code indicator; SSNI, subsystem number indicator; GTI, global title indicator; RTI, routing indicator. (From Rec. Q.713. Courtesy of ITU-T.)

Par.1 Called Party Address (CDA). This variable-length parameter is a composite, consisting of several subparameters. It can contain any combination of point code (PC), subsystem number (SSN), and/or global title (GT). The general format is shown in Fig. 15.2-1.

Octet 1 indicates the presence or absence of PC, SSN, and GT in the address.

Point Code Indicator (PCI). This indicates the presence (1) or absence (0) of PC.

Subsystem Number indicator (SSNI). This indicates the presence or absence of SSN.

Global Title Indicator (GTI). If GTI = 0000, no global title is present. If GTI is 0001, 0010, 0011, or 0100, a GT is present. The GT format depends on the value of GTI.

Routing Indicator (RTI). If the called address includes a global title, the sending SCCP specifies with RTI how the receiving SCCP should route the message:

RTI = 0 Global title translation should be performed, and routing should be based on the translation result

RTI = 1 GT translation should not be performed, and routing should be based on the destination point code (DPC) in the MTP routing label of the message (Section 8.8)

Octets 2 through n contain PC, SSN, and GT, if included.

Point Code (PC). The point codes in ITU-T SS7 signaling and ANSI No. 7 signaling have, respectively, 14 and 24 bits (Section 7.2) and occupy three and two octets, respectively.

Subsystem Number (SSN). The one-octet SSN field identifies the SCCP user (subsystem). Examples of SSN codes standardized by ITU-T are:

							← **Bits**	
Subsystem	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**
SSN field not used	0	0	0	0	0	0	0	0
SCCP management	0	0	0	0	0	0	0	1
ISDN user part	0	0	0	0	0	0	1	1
Operation, maintenance, and administration part (OMAP)	0	0	0	0	0	1	0	0
Mobile application part (MAP)	0	0	0	0	0	1	0	1

Global Title. The GT contents consist of the *translation type* (TT) and the *global title address* (GTA)—see Fig. 15.2-2. We first consider GTA, which contains several parameters. The presence (P) or absence (A) of the octets with individual parameters is indicated by the value of GTI.

GT always includes GTA—consisting of a digit string (octets d through *n*) and may include the following parameters:

			← **Bits**	
Encoding Scheme (Octet b)	**4**	**3**	**2**	**1**
BCD (binary coded decimal), odd number of digits	0	0	0	1
BCD, even number of digits	0	0	1	0

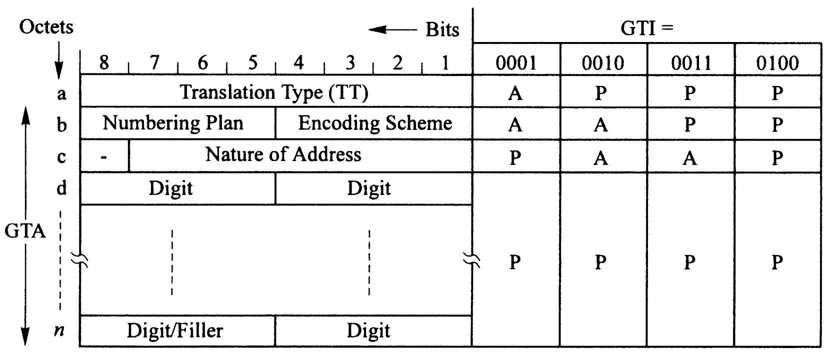


Figure 15.2-2. Global title format. P, Octet is present; A, octet is absent. (From Rec. Q.713. Courtesy of ITU-T.)

Numbering Plan (Octet b)	8	7	6	← Bits 5
Telephony/ISDN numbering plan	0	0	0	1
Maritime mobile numbering plan	0	1	0	1
Land mobile numbering plan	0	1	1	0

Nature of Address (Octet c)	7	6	5	4	3	2	← Bits 1
Subscriber number	0	0	0	0	0	0	1
National number	0	0	0	0	0	1	1
International number	0	0	0	0	1	0	0

If a GTA parameter is absent, the corresponding GT characteristic is implied. As an example, GTI is usually set to "0010" in U.S. networks. This implies that GTA is a "national number" (10 digits), that the digits are BCD coded, and that the numbering plan is the North American Telephony/ISDN numbering plan.

Translation Type (Octet a)
Most GT translations yield the PC + SSN (point code and subsystem number) of a called subsystem. TT specifies the type of translation to be performed (see Section 15.3.3). The TT values in a particular network are established by the network operator.

Par.2 Calling Party Address (CGA). The format and coding are as in Par.l. PC and SSN are always included; GT is present only in transactions between signaling points in different networks.

Par.3 Destination Local Reference (DLR). This is a reference number (three octets) that identifies a signaling connection at the SCCP in the "called" (destination) signaling point of a connection-oriented message.

Par.4 Protocol Class (PRC). This represents the class of SCCP operation (Section 15.1.2), and occupies bits 4, . . . , 1 of one octet:

Protocol Class	4	3	2	← Bits 1
Class 0	0	0	0	0
Class 1	0	0	0	1
Class 2	0	0	1	0
Class 3	0	0	1	1

Par.5 Refusal Cause. This one-octet parameter appears in the Connection Refused (CREF) message only. It indicates the reason why a requested connection cannot be set up, for example:

Refusal Cause	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**
							← Bits	
(a) Refused by called subsystem	0	0	0	0	0	0	1	1
(b) Unknown destination address	0	0	0	0	0	1	0	0
(b) Destination inaccessible	0	0	0	0	0	1	0	1
(b) Subsystem failure	0	0	0	0	1	0	1	0

Cause (a) indicates that the called subsystem has refused the connection. Causes (b) indicate that a SCCP along the message path was unable to transfer the connection request message.

Par.6 Release Cause. This is a one-octet parameter that appears in Released (RLSD) messages only. It indicates why a signaling connection is being released, for example:

Release Cause	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**
							← Bits	
(a) Originated by subsystem	0	0	0	0	0	0	1	1
(b) Subsystem failure	0	0	0	0	1	0	0	0
(b) Network failure	0	0	0	0	1	0	1	0

Cause (a) indicates a normal release of a connection, by one of the subsystems. Causes (b) originate at a SCCP and indicate that the connection has been released because of a failure.

Par.7 Return Cause. This is a one-octet parameter that appears in Unitdata Service (UDTS) messages only. It indicates why a unitdata message is being returned, for example:

Return Cause	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**
							← Bits	
(a) No translation for an address of this nature	0	0	0	0	0	0	0	0
(b) No translation for this specific address	0	0	0	0	0	0	0	1
(c) Failure of called subsystem	0	0	0	0	0	0	1	1
(d) Subsystem not equipped	0	0	0	0	0	1	0	0
(e) Signaling network failure	0	0	0	0	0	1	0	1
(f) Other failures	0	0	0	0	0	1	1	1

Return causes (a) and (b) appear in UDTS messages originated at the SCCP in a signal transfer point that has been unable to translate a global title. UDTS messages with return causes (c) and (d) are sent by the SCCP in the destination signaling point. UDTS messages, with return causes (e) and (f), can be sent by any SCCP in a signaling point along the message path.

Par.8 Return Option (RO). This appears in Unitdata (UDT) messages only. It indicates whether the UDT message should be returned to the originating subsystem if it cannot be delivered to its destination:

Return Option	**8**	**7**	**6**	**← Bits 5**
No return required	0	0	0	0
Return required	1	0	0	0

Par.9 Segmenting/Reassembling (One Octet). This appears in data form 1 messages only (class 2 connection-oriented service). Only bit 1 is used:

Bit 1	
0	Last data form of transaction
1	More data forms follow

Par.10 Sequencing/Segmenting (Fig. 15.2-3). This appears in Data Form 2 messages (class 3 connection-oriented service) only. P(S) and P(R) are send and receive sequence numbers. Bit M indicates whether more data forms follow:

M = 0	Last data form in a sequence
M = 1	More data forms follow

Par.11 Source Local Reference (SLR). This is a reference number (three octets) that identifies a signaling connection at the SCCP in the "calling" (source) signaling point of a connection-oriented message.

Par.12 Subsystem Data. This is a variable-length parameter that passes information originated by—and destined to—a subsystem. The data are transferred transparently by SCCP.

Octets ↓	8 ---------- 2	← Bits 1
1	P(S)	
2	P(R)	M

Figure 15.2-3. Sequencing/segmenting (Par.10).

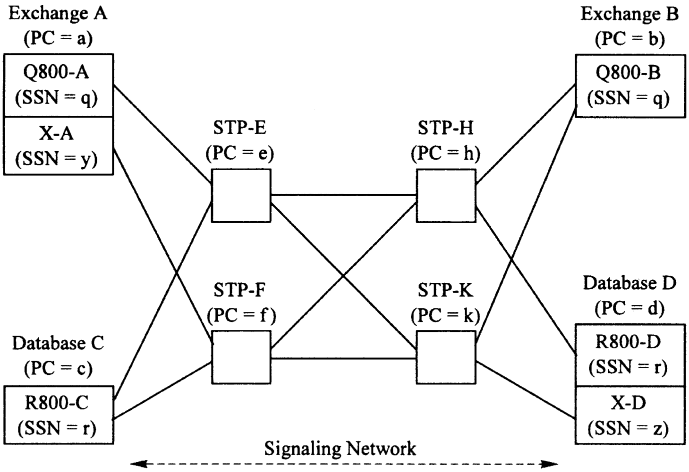


Figure 15.3-1. Transaction to obtain 800-number translation. STP, Signal transfer point.

15.3 CONNECTIONLESS SCCP

This section examines the role of connectionless SCCP in the transfer of messages between subsystems. The material in this section is illustrated by a transaction in which an exchange queries a network database to obtain the translation of an 800 number into a routing number (Section 2.1.3).

In Fig. 15.3-1, the subsystems (ASEs) denoted by Q800 and located at exchanges A and B make the 800-number queries. The subsystems denoted by R800 are located at network databases C and D and respond to these queries with messages that contain the national number of the called party and routing and charging information. Databases C and D are "mates" and contain the same information. Normally, C and D are both in-service, and the network sends 50% of the query traffic to each of them. However, if C or D is out-of-service, all queries are sent to its mate.

When a R800 has received a query, it sends a response message with routing instructions for the 800 call.

Points E, F, H, and K are signal transfer points in the signaling network. The point codes (PCs) in signaling points A through K are a through k. The subsystem numbers (SSNs) for the Q800 and R800 ASEs are q and r.

When reading the material that follows, it is helpful to look up the parameter descriptions in Section 15.2.3.

15.3.1 Unitdata Messages

In connectionless service, transaction messages are passed between SCCP and MTP as SCCP Unitdata messages—see Fig. 15.3-2 [5]. The message type code (MTYP) indicates Unitdata message. All message parameters are mandatory.

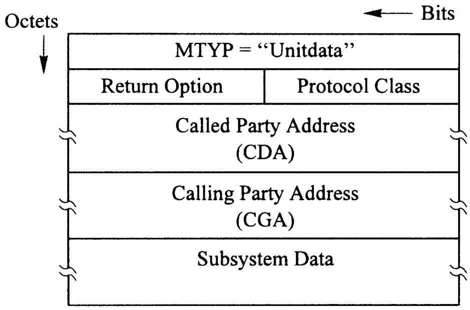


Figure 15.3-2. SCCP Unitdata message. (From Rec. Q.713. Courtesy of ITU/T.)

Par.1 Called Party Address (CDA). This is the address of the subsystem for which the message is intended.

Par.2 Calling Party Address (CGA). This is the address of the subsystem that originated the message.

Par.4 Protocol Class (PRC). This indicates the requested service class: basic connectionless service (0) or connectionless service with sequence control (1).

Par.8 Return Option (RO). This indicates whether the calling subsystem wishes to be informed if its message cannot be delivered.

Par.12 Subsystem Data. This can be a TCAP message (package) or an ISUP message.

15.3.2 Primitives

In connectionless SCCP, subsystem data are passed in N-unitdata requests and indications. The parameters in the primitives are shown in Fig. 15.3-3 [4]. On receipt of an N-unitdata request, SCCP forms a Unitdata message, using the information in the request. In connectionless service, SCCP does not maintain address information for the messages of individual transactions, and all N-unitdata requests therefore include a called and calling party address.

SCCP forms the Unitdata message. SCCP routing control (Section 15.1.3) determines the destination point code (DPC) and signaling link selector (SLS) for the message and passes the outgoing Unitdata message to MTP in a MTP-transfer request (see Fig. 15.1-2 and Section 8.8). Service indicator (SI) value = 0011 indicates a SCCP message.

Incoming messages with SI = 0011 are passed by MTP to SCCP, in MTP-transfer indications. SCCP routing control passes the subsystem data in

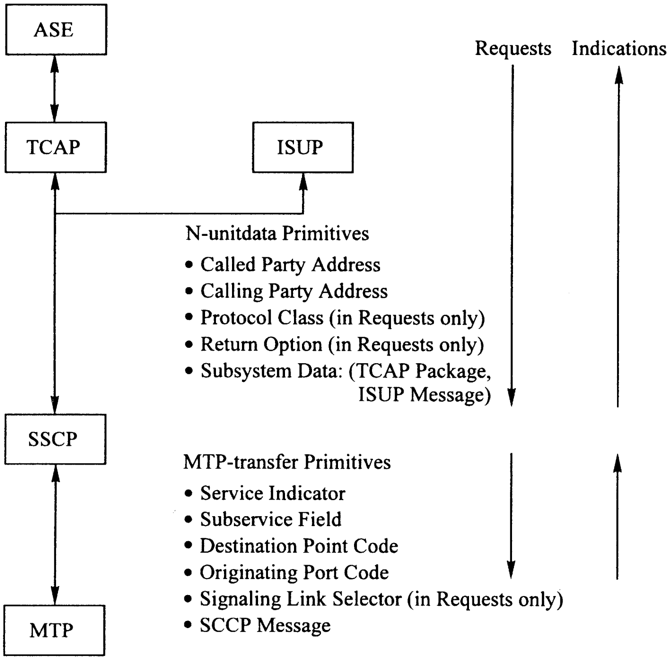


Figure 15.3-3. N- and MTP-primitives. (From Rec. Q.711. Courtesy of ITU-T.)

N-unitdata indications to ISUP or TCAP, depending on the value of SSN in the called party address.

The N-unitdata indications include all information listed in Fig. 15.3-3, except the protocol class and return option (they are not significant to the receiving TCAP or ISUP). TCAP then delivers the information in the package to the subsystem specified by SSN.

15.3.3 Global Title Translation

A SCCP called party address (Par.1) can contain various combinations of PC, SSN, and/or GT. We now explore the reasons for global titles and GT translations.

A subsystem in a network is uniquely identified by the combination of the point code PC of its signaling point and its SSN at that point. If a SCCP message has a PC + SSN called party address, the MTPs along the message path then use PC to route the message to its destination signaling point, and the SCCP at this point delivers it to the subsystem specified by SSN. For example, in Fig. 15.3-1, a SCCP message with called party address PC = c, SSN = r is delivered to subsystem R800 in database C.

A global title is a "functional" address of a subsystem in an exchange or database. The reason for these functional addresses is illustrated with the example of Fig. 15.3-1.

Suppose that the setup of a call with a called 800 number, say, 800-123-4567, has reached exchange A. Subsystem Q800-A then has to send a query message to an R800 subsystem at a database that stores the translation for this number—in this example, databases C and D.

In principle, exchange A could translate the received 800 number into a PC + SSN called party address of an R800 in the appropriate database. However, this would require "800-number" translation tables in all exchanges and would entail a large effort when an item in the tables has to be added or removed.

A better arrangement is to let the exchanges use the received 800 number as a global title (GT) address, which is the "functional" address of an R800 at a database with information on the 800 number, and install the GT translation capability in the SCCPs of the signal transfer points (STPs) of the network. The SCCP at the originating exchange then routes the query message to a directly connected STP, whose SCCP translates GT into the PC + SSN address of the destination. From this point on, the message can be routed by MTP to its destination.

The number of exchanges in a network greatly exceeds the number of STPs, and placing the GT translation capability in the SCCPs at the STPs instead of in the exchanges greatly reduces the effort to update the translation data.

A GT consists of two parts, known as the GT address (GTA), which is a digit string received from a calling subscriber, and the translation type (TT), which indicates the desired translation (Fig. 15.2-2). For example, if GTA is the national number of a subscriber (S), one value of TT could indicate that GTA is to be translated into the PC + SSN address of an ASE in the maintenance center that covers S; another TT value could require a translation that yields the PC + SSN address of an ASE in the revenue accounting center that covers S, and so on.

15.3.4 Transfer of Unitdata Messages

We now examine the transfer of Unitdata messages for an 800-number query—response transaction—see Fig. 15.3-1. We assume that all entities in the figure are part of the same telecommunication network.

Query Message. Suppose that the setup of an 800 call has reached exchange A. The Q800-A then launches a query to a database, to obtain the routing number for the call. We assume that the call-routing information for this particular 800 number is stored in R800-C and R800-D (at databases C and D).

Figure 15.3-4 shows the MTP and SCCP address parameters in the primitives at signaling points A, E, and D and in the message signal unit (MSU) that carries the first unitdata message.

TCAP-A has received the calling and called address (CDA, CGA) from Q800-A and includes them in the N-unitdata request to SCCP-A. The CDA is a GT in which the translation type = t and the address = n (the called 800 number). The translation type indicates that n has to be translated into the SP + SSN address of an R800 at a database with information on that number.

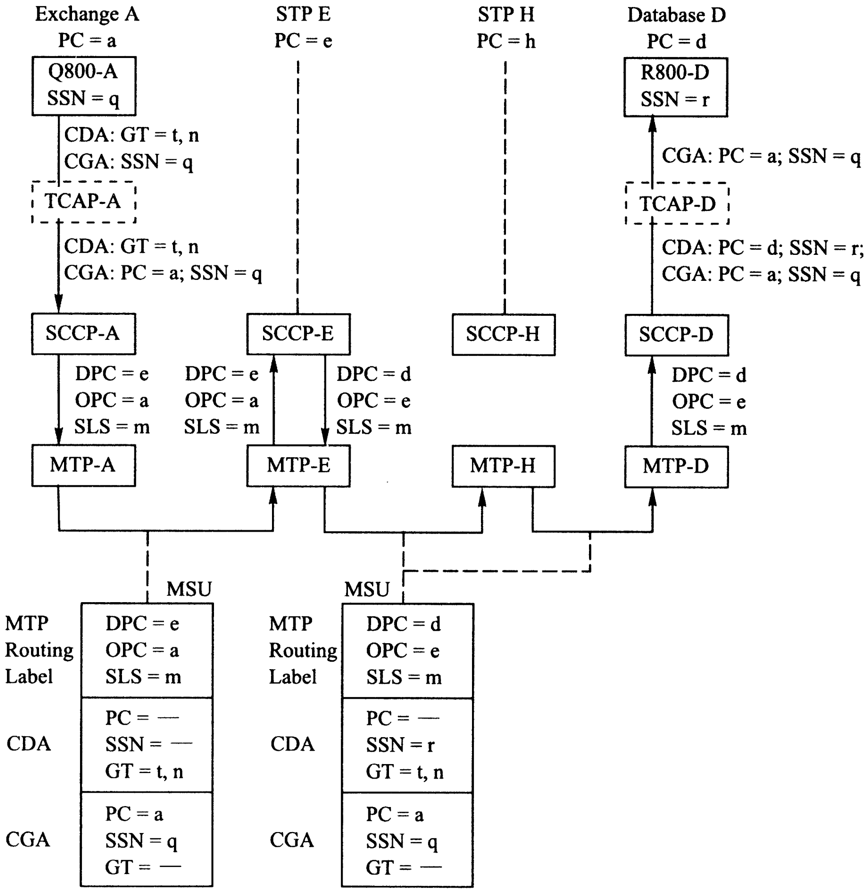


Figure 15.3-4. Address information in the first SCCP Unitdata message of a transaction.

SCCP-A enters GT in the CDA field of the Unitdata message, PC = a, SSN = q in the CGA field, and sets the routing indicator (Fig. 15.2-1) to RTI = 0, indicating that a GT translation is needed. Since SCCP-A has received a GT called address and knows that the SCCP-E can perform GT translations, it includes the MTP address of signaling point E (DPC = e) when passing the message. MTP-A then forms the MSU and transfers it to STP-E.

SCCP-E translates the GT and obtains the addresses PC = c, SSN = r and PC = d, SSN = r of the R800 units at databases C and D. Assuming that SCCP-E selects database D, it enters SSN = r in the CDA of its outgoing message and passes it to its MTP-E, in a MTP-transfer request that includes destination point code DPC = d. It also sets the routing indicator to RTI = 1. The MTP-E routes the MSU to MTP-H, which routes it to database D.

SCCP-D receives the message from MTP-D and passes it to TCAP-D in an N-unitdata indication, which includes the calling address PC = a, SSN = q. Finally, TCAP-D delivers the message to R800-D.

Response Message. R800-D uses the received CGA (PC = a, SSN = q) as the CDA for its response message. SCCP-D passes the message to MTP-D, in a MTP-transfer request that includes DPC = a. The transfer of the message to SCCP-A is done by MTPs exclusively (no GT translation needed). At exchange A, subsystem number SSN = q in the CDA is used to deliver CGA and the routing number for the call (in the subsystem data field) to Q800-A. Exchange A then routes the call to its destination.

The transaction between Q800-A and R800-D requires just two messages. This is the case for most transactions. However, after the second message, both R800-D and Q800-A know each other's PC + SSN address. If a transaction requires additional messages, the called addresses in these messages are always PC + SSN addresses and can be transferred by MTPs along the signaling route (no GT translation required).

15.3.5 Final and Intermediate GT Translations

In the above example, SCCP-E makes a *final* GT translation, which yields the PC + SSN address of destination R800-D. This is possible because a SCCP in a signal transfer point has the necessary data to perform final translations for all destinations in its network, and we have assumed that all entities in Fig. 15.3-4 are in the same network.

When the originating and destination ASEs are in different networks, the SCCPs in the originating network have no data to do final GT translations. In this case, the message transfer involves one or more *intermediate* GT translations to route the message to a STP in the destination network, and a final GT translation by the SCCP at that STP.

Figure 15.3-5 shows an example. We consider long-distance networks in countries 1 and 2 and explore the transfer of an initial message by originating ASE-A to destination. ASE-E.

Each national network has *international* STPs (STPI), whose SCCPs can do final GT translations for destinations in their respective countries, and intermediate GT translations to route a message to a STPI in the destination country.

In the figure, the point codes of signaling points in countries 1 and 2 are denoted by la, lb, 1c, and 2d, 2e. The STPIs also have an international point code (ic, id).

We now examine the routing of the first Unitdata message of an international calling-card verification transaction. An operator at exchange A has received a call from a caller who has a calling card issued in country 2 and wants the call to be charged to the card. We denote the calling card number by $2n$. Before extending the call setup, the operator needs to verify the validity of the calling card. Information about the validity of $2n$ is available only in country 2 (at ASE-E).

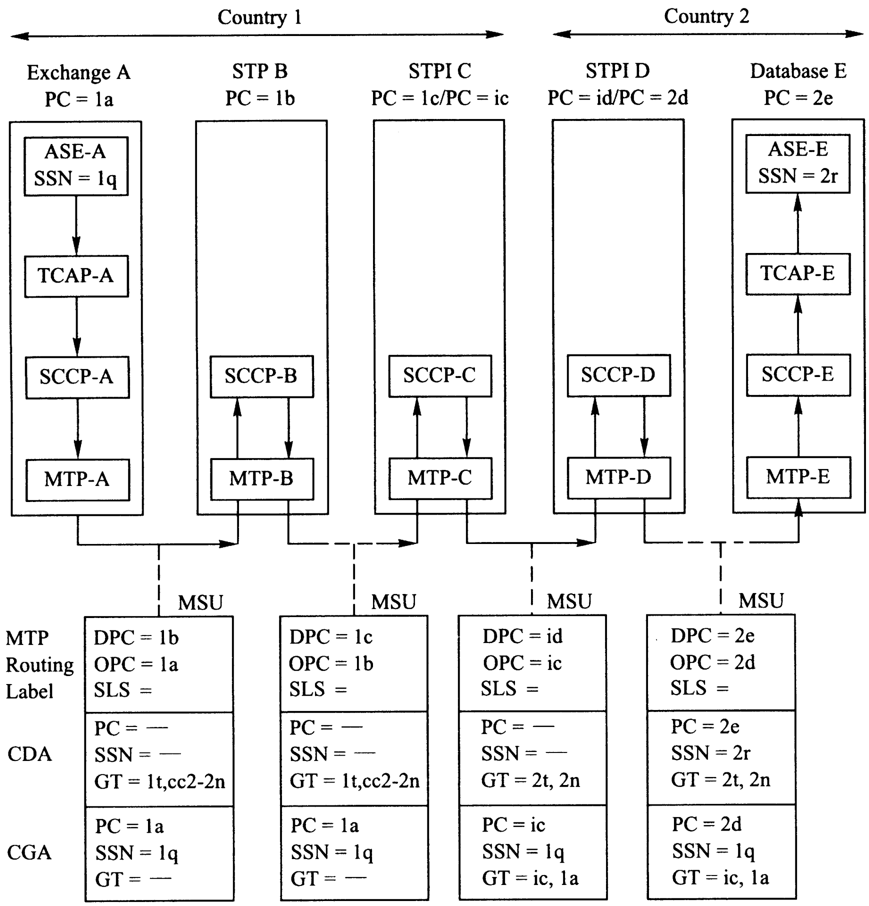


Figure 15.3-5. Address information in the first SCCP Unitdata message of an international transaction.

By entering the country code (cc2) of country 2 and $2n$, the operator causes ASE-A to start the query. The ASE generates the global title address GTA = cc2-$2n$ and translation type TT = 1t (indicating international calling-card verification). SCCP-A sets RTI to 0 and routes the message to STP B.

SCCP-B performs an intermediate translation (based on 1t and cc2) which yields the point code (DPC = 1c) of STPI-C. The message reaches SCCP-C, which does another intermediate translation, based on 1t, cc2, and possibly the initial digits of $2n$. This yields the international point code (id) of signal transfer point D and translation type 2t (the value used in country 2 for calling-card verification). SCCP-C also removes the country code from the GTA.

SCCP-D does the final GT translation—based on 2t and $2n$—which yields the PC + SSN address of ASE-E.

Since SCCP-B and SCCP-C have done intermediate translations, their outgoing messages include RTI = 0. SCCP-D has done a final translation and includes RTI = 1.

Note that SCCP-C enters ic (its international point code) and 1a (the point code of A) into the calling (CGA) global title field of its message. The N-unitdata indication passed from SCCP-E to ASE-E includes the CGA address parameters: 1a, 1c, and 1q (SSN of ASE-A), and 2d (the point code of D).

When responding, ASE-E passes the CDA global title parameters 1a, 1c, and 1q, along with the appropriate translation type and the destination point code, DPC = 2d, to SCCP-E.

The intermediate GT translations at D and C then simply consist of using 1c and 1a as destination point codes for their outgoing unitdata messages.

15.3.6 Class 0 and Class 1 Service

A calling subsystem requests class 0 or class 1 service with Par.4 (protocol class). In class 0 (basic connectionless) service, the calling SCCP is free to assign any value of SLS (signaling link selector) to its outgoing messages. This means that two consecutive outgoing messages from a subsystem may traverse different routes in the signaling network and could arrive out-of-sequence (Section 8.8.5).

In class 1 (sequence-controlled connectionless) service, SCCR assigns the same SLS value to all outgoing messages of a particular subsystem. Consecutive outgoing messages of a subsystem then traverse the same links in the signaling network. The signaling links themselves maintain in-sequence delivery of MSUs, even when transmission errors occur (Section 8.3).

One way to associate the same SLS value to all outgoing messages of a subsystem is to use the low-order bits of the calling subsystem's SSN as SLS.

15.3.7 Unitdata Service Message

A SCCP in a signaling point along the path of a Unitdata message may determine that it cannot transfer a received Unitdata message to its destination. For example, a SCCP in a STP may not be able to translate the global title address of the called party, or the SCCP at the destination signaling point may find that there is no subsystem at the signaling point that corresponds to the SSN in the received message, or that the called subsystem is out of service.

If Par.8 (return option) in a Unitdata message that cannot be delivered indicates that the message should be returned, the SCCP sends a Unitdata Service Message (UDTS) to the calling SCCP. This message includes the subsystem data of the received message and a Par.7 (return cause) that indicates why the message is being returned. On receipt of a UDTS message, SCCP alerts the calling subsystem with a N-notice indication (Fig. 15.1-2) that includes the called address, the subsystem data, and the return cause.

15.4 CONNECTION-ORIENTED SCCP

In this mode of operation, a (virtual) connection is set up before data transfer between two SCCP users takes place. We distinguish two connection types. *Temporary* connections are established and released at the start and end of a transaction. *Permanent* connections are long-term connections that can be set up and released only by administrative or maintenance personnel.

Connection-oriented service is the preferred way for transactions that involve the transfer of large amounts of data, which puts a momentary heavy load on the involved subsystems and the signaling network. Transactions of this nature are not call-related and can be deferred, say, for several minutes. Connection-oriented service gives the called subsystem an opportunity—at the time it receives a connection request—to determine whether it can handle the transaction at this time. If yes, it accepts the request. If not, it refuses. This avoids cluttering up the network with messages that cannot be processed anyway.

ITU-T has defined two classes of connection-oriented service [4,7]: *basic connection-oriented service* (class 2) and *connection-oriented service with flow control* (class 3).

When reading this section, it is helpful to look up the message contents (Table 15.2-2), and the parameter descriptions (Section 15.2.3).

15.4.1 Primitives

The primitives and connection-oriented service are shown in Fig. 15.1-2. The N-connect and N-disconnect primitives are used to establish and release connections. The N-data primitives are used for the transfer of user data during the connection.

In addition to the familiar *request* and *indication* primitives, SCCP also uses *response* and *confirmation* primitives. They play a role in the establishment of a connection.

15.4.2 Connection-Oriented Class 2 Service

Figure 15.4-1 shows the SCCP messages and primitives during a temporary connection with class 2 service. The transaction involves subsystem P at signaling point A, and subsystem Q at signaling point B.

Establishing the Connection. Subsystem P initiates the transaction and passes an N-connect request to SCCP-A. The SCCP sends a Connection Request message to SCCP-B, which passes an N-connect indication to subsystem Q. The subsystem decides to accept the connection request and passes an N-connect response to SCCP-B, which now sends a Connection Confirm message. SCCP-A informs subsystem P that the connection request has been accepted, with an N-connect confirmation.

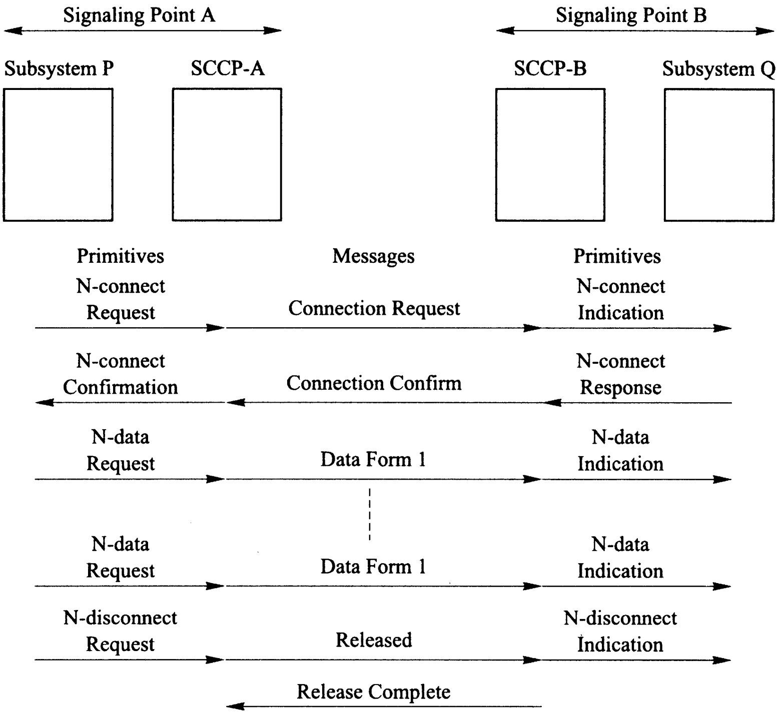


Figure 15.4-1. Primitives and messages in a class 2 connection. (From Rec. Q.711. Courtesy of ITU-T.)

Data Transfer. The transfer of subsystem data begins. Subsystem P passes its data to SCCP-A in an N-data request. SCCP-A places the data in Par.12 (subsystem data) of a Data Form 1 message and sends it to SCCP-B. The latter SCCP extracts the subsystem data and passes the data to subsystem Q in an N-data indication. Figure 15.4-1 shows data transfer in one direction only. However, data transfer in both directions is allowed.

Release of Connection. The connection can be released by either subsystem. In this example, P initiates the release, passing an N-disconnect request. SCCP-A releases the connection at its end and sends a Released message. SCCP-B then passes an N-disconnect indication to subsystem Q, releases the connection at its end, and sends a release complete message.

15.4.3 Records and References

During the exchange of the Connection Request and the Connection Confirm messages, the subsystems and SCCPs allocate reference numbers to identify the connection and build records that store reference numbers and address parameters that are associated with the connection. The records are consulted when sending—or receiving—a data form message and are discarded at the end of the connection. The reference numbers are of the following types.

Connection Identifier CID. Subsystems and SCCPs can be involved with several simultaneous connections. A connection identifier uniquely identifies a connection at a signaling point. It is stored in the records of the subsystem and SCCP and is included in the N-primitives between them.

Source Local Reference (SLR, Par.11). This identifies the connection, as known by the SCCP. It is included in SCCP messages to and from the other SCCP.

Destination Local Reference (DLR, Par.3). This identifies the connection, as known by the SCCP at the other end of the connection, and is included in SCCP messages.

Allocation of Connection Identifiers and Source Local References. The SCCP at each signaling point has a pool of available CIDs and a pool of available SLRs. A subsystem, or a SCCP, can seize an available CID or SLR from a pool and allocate it to the new connection by entering it into its connection record. At the end of the connection, CID and SLR are returned to their respective pools.

The records for the connection of Fig. 15.4-1 are shown in Fig. 15.4-2. A SCCP record is accessed with parameter CID or SLR. When a record has been accessed, its other parameters become available.

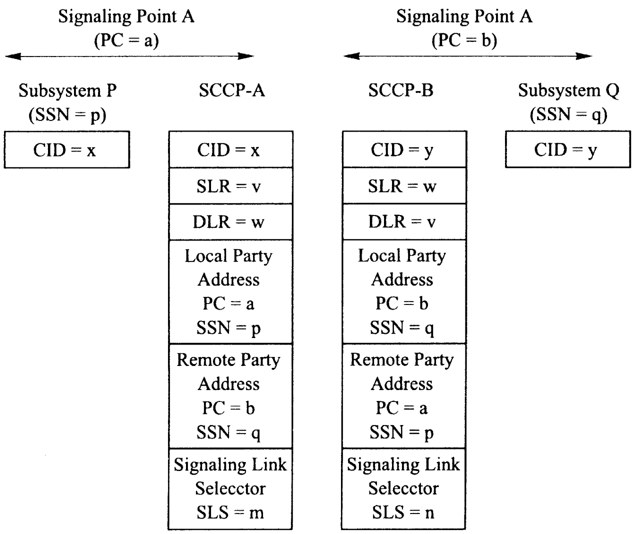


Figure 15.4-2. Connection records.

Building the Records. We explore the buildup of the records. Subsystem Q initiates the transaction. It establishes a connection record and allocates the connection identification value (CID) = x. It then passes an N-connect request that includes the called and calling party addresses, and CID. The called party address may be a global title or a DPC + SSN address. The calling party address is always in the format PC + SSN (PC = a, SSN = p).

On receipt of the request, SCCP-A establishes a connection record. It enters the received connection identifier (CID = x) in the record. The calling party address (PC = a, SSN = p) is entered as the "local" party address. SCCP-A then allocates a source local reference value (SLR = v) and a signaling link selector value (SLS = m) and sends a Connection Request message that includes the called and calling party addresses (Par.l and Par.3) and the source local reference (Par.11).

On receipt of the message, SCCP-B establishes its record. It enters the received called party address (which may have undergone global title translation and now is PC = b, SSN = q) as the "local" party address. The received calling party address (PC = a, SSN = p) and source local reference (SLR = v) are entered as the "remote" party address and "destination" local reference (DLR). SCCP-B also allocates a connection identifier CID = y. It then passes an N-connect indication to subsystem Q (the "local" party) that includes the connection identifier.

Assuming that the subsystem accepts the connection request, it builds a connection record and enters the received connection identifier (CID = y). It then passes an N-connect response to SCCP-B, including CID = y.

SCCP-B accesses the record in which CID = y and allocates a source local reference (SLR = w) and a signaling link selector (SLS = n). It then sends a Connection Confirm message, using the "remote" party address (PC = a, SSN = p) as called party address, the "local" party address (PC = b, SSN = q) as calling party address (PC = b, SSN = q), the source local reference (SLR = w), and the destination local reference (DLR = v).

The SCCP message arrives at SCCP-A. It accesses the record whose source local reference matches the received DLR = v. It then enters the received calling party address as "remote" party address and passes an N-connect confirmation that includes CID = x to the "local" party identified by SSN = p. At this point all connection records are complete, and the transfer of subsystem data can start.

15.4.4 Transfer of Subsystem Data

Figure 15.4-3 shows the transfer of a Data Form 1 message on the established connection. Subsystem P passes an N-data request that includes CID = x and the subsystem data. SCCP-A locates the record associated with the connection (which has CID = x) and determines the values of the destination local reference (DLR = w) and the "remote" point code (PC = b). It then creates a data form that includes the DLR and the subsystem data and passes the form to MTP, in a MTP-transfer request that includes the destination point code DPC = b.

When SCCP-B receives the message, it locates the record whose source local reference matches the received DLR = w. It then obtains the parameters

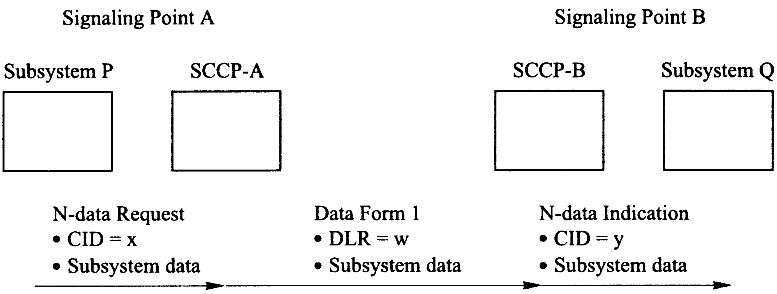


Figure 15.4-3. Transfer of a data form 1.

SSN = q, CTD = y and passes an N-data indication to subsystem Q, which includes the CID and the subsystem data.

15.4.5 Connection-Oriented SCCP Services

SCCP class 2 service provide the following services to its subsystems.

Connection Refusal. A subsystem may already be involved in one or more connections when it receives a new N-connect indication. If the already existing connections involve large data transfer activities, the subsystem can decide to refuse the new connection at this time. It does so by responding with an N-disconnect request. The local SCCP then sends a Connection Refused message to the SCCP of the subsystem that requested the connection, which is then informed by an N-disconnect indication.

In-sequence Delivery of DT1 Messages. Each SCCP allocates a SLS (signaling link selector) value when it establishes its connection record and uses this value for all outgoing messages of the connection. In each direction, all messages then follow a fixed signaling route and are delivered in-sequence by MTP (Section 8.8.5).

Segmentation and Reassembly of User Data. The maximum length of the signaling information field in a message signal unit is 272 octets (Fig. 8.8-1). After making allowances for the routing label and for the overhead in a SCCP message, about 255 octets are available for subsystem data.

Class 2 service allows a calling subsystem to pass a much larger amount of data in one N-data request. The calling SCCP then segments the data into units of suitable size and transmits them in successive DT1 messages. The called SCCP reassembles the data units and passes them in one N-data indication to the called subsystem.

This service depends on the in-sequence delivery of the DT1 messages. Each message includes a Par.9 (segmenting/reassembling). The calling SCCP sets Par.9 = 1 in all DT1 (Data Form 1) messages, except the last one. The called SCCP reassembles the data units of the consecutive DT1s until it has received a message with Par.9 = 0. It then passes the reassembled subsystem data to the called subsystem, in one N-data indication.

Class 3 Services. These include the class 2 services and *flow control*. This protects the called subsystem in a connection against being overloaded with incoming DT2 messages [3,7]. The calling SCCP assigns cyclically increasing (. . . 0, 1, . . . , 126, 127, 0, . . .) *send sequence numbers* [P(S)] to consecutive outgoing DT2 (data form 2) messages of a connection. The called SCCP sends Data Acknowledgment messages, which include a *receive sequence number* [P(R)] that represents the highest numbered (mod 128) accepted DT2. Parameters P(S) and P(R) are part of Par.10, which is included in DT2 and acknowledgment messages.

The calling SCCP may send additional DT2s as long as $P(S) \leq P(R) + W$, where P(R) is the receive sequence number in the most recently received acknowledgment message, and window W represents a fixed number of DT2 messages. The value of W is negotiated during the establishment of a class 3 connection. In this way, the called SCCP controls the message flow.

15.5 SCCP MANAGEMENT

The purpose of SCCP management (SCMG) is to maintain—if possible—the transfer of SCCP messages when failures occur in the signaling network and/or subsystems, and to inform SCCP users to stop sending messages that cannot be delivered. If a failed subsystem has a duplicate, SCMG requests its SCCP to reroute messages to the backup subsystem [6].

15.5.1 SCMG Interfaces

SCMG has interfaces with the subsystems, the MTP, and the SCCP connectionless control at its signaling points—see Fig. 15.1-2.

N-Primitives are the interface with the local subsystems. The subsystems pass information, to SCMG in N-requests and N-responses, and SCMG passes information to the subsystems in N-indications and N-confirmations.

The interface with SCCP connectionless control allows SCMG to send and receive SCCP Unitdata messages to/from SCMGs in other signaling points.

The interface with the local MTP consists of MTP-pause, MTP-resume, and MTP-status indications (Section 8.9.1).

15.5.2 N-Primitives and Their Parameters

Primitives. The N-primitives (Fig. 15.1-2) are outlined below:

N-State Primitives. These primitives indicate a change in status of a subsystem. A subsystem indicates a status change to its SCMG in an N-state request, and

TABLE 15.5-1 SCCP Management Primitives and Parameters

	Parameter Acronyms				
Primitives	AFSN	AFPC	US	SPS	SMI
N-State					
Request	X		X		
Indication	X	X	X		X
N-PC-State					
Indication		X		X	
N-Coord					
Request	X				
Indication	X	X			X
Response	X				
Confirmation	X	X			X

Note: AFSN, affected subsystem number; AFPC, affected point code; US, user (subsystem) status; SPS, signaling point status; SMI, subsystem multiplicity indicator.

Source: Rec. Q.711. Courtesy of ITU-T.

SCMG informs its local subsystems about a status change of a subsystem at another signaling point with an N-state indication.

N-PC-State Indications. These primitives are passed by SCMG to its local subsystems to indicate the status of a signaling point.

N-Coord Primitives. These primitives are used when a subsystem that has a backup subsystem (say, a network database) wants to go out of service.

Parameters. The parameters in N-primitives are listed in Table 15.5-1.

Subsystem Status. This indicates the new status of a subsystem (“in-service” or “out-of-service”).

Affected Subsystem. This identifies the subsystem whose status has changed, by subsystem number SSN.

Subsystem Multiplicity. This indicates whether the affected subsystem has a duplicate copy in the network.

Affected Point Code. This is the point code of a signaling point whose status has changed, or where a subsystem has changed status.

Signaling Point Status. This indicates the new status of a signaling point (“inaccessible” or “accessible”).

15.5.3 SCMG Messages and Parameters

The SCMGs at different signaling points use the connectionless services of SCCP to send Unitdata messages to each other. A SCMG is thus both a part and a user of its SCCP. A similar situation exists in the signaling network management part of MTP (Section 8.8). ITU-T has specified SSN = 0000 0001 as the subsystem number of SCMG.

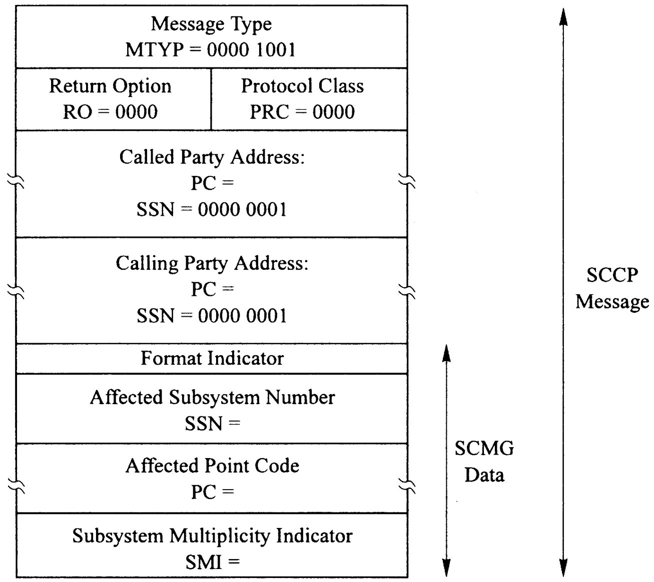


Figure 15.5-1. SCMG message. (From Rec. Q.712. Courtesy of ITU-T.)

The contents of SCMG messages are shown in Fig. 15.5-l. Message type code (MTYP = 0000 1001) identifies the message as a Unitdata message. Protocol class and return option are coded PRC = 0000 (basic connectionless service) and RO = 0000 (no return of messages that cannot be delivered). Subsystem numbers (SSNs) in the called and calling party address are coded 0000 0001. This identifies the message as a SCMG message. The *format indicator* identifies a particular SCMG message type:

Subsystem Allowed. The subsystem can be used.

Subsystem Prohibited. The subsystem cannot be used.

Subsystem Status Test. A request for subsystem status information.

Subsystem Out-of-Service Request. A request by a subsystem to go out-of-service.

Subsystem Out-of-Service Grant. A permission for a subsystem to go out-of-service.

The format identifier codes are lited in Table 15.5-2. The message parameters have the same meanings as in the primitives.

15.5.4 SCMG Procedures

The N-state requests from its local subsystems, the MTP indications from the local MTP, and the received SCMG messages keep a SCMG informed on status changes

TABLE 15.5-2 Format Identifiers in SCMG Unitdata Messages

Format Identifier Name	Code
Subsystem allowed	0000 0001
Subsystem prohibited	0000 0010
Subsystem status test	0000 0011
Subsystem out-of-service request	0000 0100
Subsystem out-of-service grant	0000 0101

Source: Rec. Q.713. Courtesy of ITU-T.

of subsystems and signaling points throughout the network. A status change can trigger the execution of a SCMG procedure.

There are two groups of SCMG procedures: subsystem management and point code management. The next sections give a number of examples, using the configuration of Fig 15.3-1. The SCMGs and subsystems at the various signaling points are denoted as SCMG-A (SCMG at signaling point A), R800-D (R800 subsystem at signaling point D), and so on.

In the material that follows, a *concerned* signaling point or subsystem is a signaling point or subsystem that has to be informed immediately when the status of a subsystem or signaling point changes.

15.5.5 Examples of SCMG Procedures

Subsystem Status Test. This test is performed as a part of other procedures. In Fig. 15.3-1, when SCMG-A has received a subsystem-prohibited message for R800-C, it periodically audits the status of the subsystem, by sending subsystem-status-test messages. If SCMG-C determines that R800-C is in-service again, it responds with a subsystem-allowed message. Otherwise it does not respond.

Broadcast Procedure. This procedure is used when a status change occurs in a subsystem that has a number of concerned signaling points. Suppose that subsystem R800-D has gone out-of-service. It then passes an N-state out-of-service request to SCMG-D. The STPs in the network are concerned about R800-D, because they perform global title translations for 800-query messages that yield a primary and a backup R800 subsystem. Assume R800-D and R800-C have been designated as the primary and backup subsystems. SCMG-D now broadcasts subsystem-prohibit messages for R800-D to the STPs. The messages indicate that R800-D has a duplicate. The SCMGs at the STPs then inform their local SCRC (SCCP routing control) functions to route messages for R800-D to the backup system R800-C. The SCMGs also start subsystem status tests of R800-D.

When R800-D goes in-service again, SCMG-D broadcasts subsystem-allowed messages to the SCMGs at the STPs. When a SCMG receives the subsystem-allowed message from SCMG-D (either as a broadcast message or in response to a subsystem-status-test message), it requests its SCRC to resume routing to R800-D.

Subsystem-Prohibited Messages to Individual Subsystems. Suppose that subsystem X-D at signaling point D goes out-of-service, and that SCMG-D has determined that X-D has no concerned subsystems (Fig. 15.3-1). Therefore, subsystem-prohibited messages are not broadcast. However, when SCCP-D now receives a SCCP message for X-D from, say, X-A, it informs SCMG-D. The SCMG then sends a subsystem-prohibited message for X-D to SCMG-A. This SCMG informs X-A, by passing an N-state out-of-service indication for X-D. SCMG-A also starts a subsystem status test for subsystem X-D.

When SCMG-D determines that X-D is in-service again, it responds to the next subsystem-status-test message with a subsystem-allowed message, and SCMG-A then passes an N-state in-service indication to X-A.

Coordinated Status Changes. With this procedure, a subsystem that has a duplicate and wishes to go out-of-service for a modification or scheduled maintenance first checks whether its backup subsystem can take over the load. Consider the case that R800-D wishes to go out-of-service. It passes an N-coord request to SCMG-D, which then sends a subsystem-out-of-service request message to SCMG-C. This SCMG passes an N-coord indication to the backup subsystem R800-C.

If the subsystem agrees to take over the load, it passes an N-coord response to SCMG-C, which then sends a subsystem out-of-service-grant message to SCMG-D. This SCMG then passes an N-coord confirmation primitive to R800-D, to indicate that it can go out-of-service. It also broadcasts subsystem-prohibited messages to the concerned subsystems as described above.

If R800-C cannot take over the load of its mate, it does not respond to the N-coord indication, and no subsystem-out-of-service-grant is sent to SCMG-D. When R800-D does not receive the N-coord confirmation for its request within a certain time, it knows that its request has been denied.

15.5.6 Point Code Management Procedures

Point code management procedures are triggered by the receipt of MTP indications. Consider the receipt at SCMG-A of a MTP-pause indication for affected signaling point D in Fig. 15.3-1. SCMG-A then informs SCRC-A that destination D and its subsystems are prohibited, and SCRC-A stops routing messages to D. SCMG-A also locally broadcasts N-PC-state signaling-point-inaccessible indications for destination D, and N-state out-of-service indications for the subsystems at D, to its local subsystems.

When SCMG-A receives a MTP-resume indication for destination D, it informs SCRC-A that the destination is accessible again and broadcasts N-PC-state signaling-point-accessible indications for destination D to its local subsystems. It then updates its status information on the subsystems at D, by sending subsystem-status-test messages. On receipt of a subsystem-allowed message for a subsystem at D, it passes N-state in-service indications for that subsystem to its local subsystems.

15.6 ACRONYMS

AFPC	Affected point code
AFSN	Affected subsystem number
ANSI	American National Standards Institute
ASE	Application service element
BCD	Binary coded decimal
CC	Connection confirm message
CDA	Called address
CGA	Calling address
CID	Connection identifier
CR	Connection request message
CREF	Connection refused message
DPC	Destination point code
DLR	Destination local reference
DTI	Data form 1 message
DT2	Data form 2 message
FI	Format indicator
GSM	Global System for Mobile Communications
GT	Global title
GTA	Address information in GT
GTI	Global title indicator
ISDN	Integrated services digital network
ISUP	ISDN User Part
ITU	International Telecommunication Union
ITU-T	Telecommunication Standardization Sector of ITU
MSU	Message signal unit on signaling link
MTP	Message Transfer Part
MTYP	Message type
NSP	Network service part
OPC	Originating point code
OSI	Open Systems Interconnection
PC	Point code
PCI	Point code indicator
PRC	Protocol class
Q800	Subsystem (ASE) making 800-number queries
REF	Reference number
RLC	Release complete message
RLSD	Released message
RO	Return option
RTI	Routing indicator
SCCP	Signaling Connection Control Part
SCLC	SCCP connectionless control
SCMG	SCCP management
SCOC	SCCP connection-oriented control

SCRC	SCCP routing control
SLR	Source local reference
SLS	Signaling link selector
SMI	Subsystem multiplicity indicator
SOG	Subsystem Out-of-Service Grant message
SOR	Subsystem Out-of-Service Request message
SREF	Source reference number
SSA	Subsystem Allowed message
SSN	Subsystem number
SSNI	Subsystem number indicator
SSP	Service switching point
SSP	Subsystem Prohibited message
SST	Subsystem Test message
SS7	Signaling System No. 7
STP	Signal transfer point
TCAP	Transaction Capability Application Part
TT	Translation type
TUP	Telephone User Part
UDT	Unitdata message
UDTS	Unitdata Service message
UM	User message

15.7 REFERENCES

1. A. R. Modaressi and R. A. Skoog, Signaling System No. 7: A Tutorial, *IEEE Commun. Mag.*, **28**, No. 7, July 1990.
2. R. Manterfield, *Common Channel Signalling*, Chapter 5, Peter Peregrinus Ltd., London, 1991.
3. P. G. Clarke and C. A. Wadsworth, CCITT Signalling System No. 7: Signalling Connection Control Part, *Br. Telecommun. Eng.*, **7**, Apr. 1988.
4. *Signalling System No. 7, Funtional Description of the Signalling Connection Control Part*, Rec. Q. 711, ITU-T, Geneva, 1993.
5. *Signalling System No. 7, Definition and Functions of SCCP Messages*, Rec. Q. 712, ITU-T, Geneva, 1993.
6. *Signalling System No. 7, SCCP Formats and Codes*, Rec. Q. 713, ITU-T, Geneva, 1993.
7. *Signalling System No. 7, SCCP Procedures*, Rec. Q. 714, ITU-T, Geneva, 1993.
8. *American National Standard for Telecommunications—Signalling System No. 7 (SS7)—Signaling Connection Control Part*, ANSI T.112, ANSI, New York, 1992.

16

TRANSACTION CAPABILITIES APPLICATION PART

16.1 INTRODUCTION

The Transaction Capabilities Application Part (TCAP) of Signaling System No. 7, in conjunction with the Signaling Connection Control Part (SCCP) and the Message Transfer Part (MTP) [1–3], enables application entities (AEs) at two *nodes* (the TCAP term for *signaling points*) to conduct transactions based on remote operations. MTP and SCCP are discussed in Chapters 8 and 15.

TCAP, or TC for short, is similar in many respects to the protocols defined by ITU-T for data communication networks, which are documented in X-series recommendations.

TCAP was first specified by the CCITT in 1989 and the recommendations have been revised by ITU-T in 1993 and 1997 [4–8].

The U.S. version of TCAP has been specified by ANSI [9]. The work on that version started before the publication of the initial CCITT recommendations. As a consequence, there are pronounced differences in terminology, and some differences in coding, between the two versions. Sections 16.1–16.3 of this chapter follow the ITU-T recommendations. Section 16.4 describes the main differences with the U.S. version. The European Telecommunications Standards Institute (ETSI) has also specified its own TCAP requirements [10], mentioned in Section 16.5.

16.1.1 Transactions, Remote Operations, and Application Contexts

As explained in Chapter 14, during a transaction, one (usually) or more (rarely) remote operations are executed: the operation is requested by an application service element (ASE) belonging to an AE in one node and executed by a "peer"

AE/ASE at another node [11]. An ASE that uses TC services is referred to as a *TC-user* in ITU-T recommendations, and in the rest of this chapter we will use that term when focusing on the internal communication between a TC-user ASE and TCAP. TCAP is itself an ASE and serves all ASEs in a node.

The TCAP protocol provides communication between two nodes containing ASEs involved in transactions for remote operations. Some remote operations result in information being sent back to the requesting ASE. For example, in the "800-number calling" application, the ASE at an exchange requests its peer to translate an 800 number into a routing number. Other remote operations result in instructions sent back to the requesting ASE. In the chapters that follow we shall encounter operations of both types.

Examples of TC-user ASEs are intelligent network (IN) services of the Intelligent Network Application Part (INAP) and Mobile Application Part (MAP) services. MAP and INAP are described in the next chapters, where examples of TCAP usage and contents are provided.

Remote operations are handled by TCAP information elements (IEs) called *components*, which are contained in the component portion of the message and follow the requirements of the ROSE protocol (Chapter 14) [11–19]. TCAP incorporates ROSE directly (with minor additions), without a ROSE layer/sublayer.

In addition to handling operation-carrying messages, TCAP allows two AEs to negotiate the *application context* (AC) for their communication, using information in the dialog control portion of the message. An application context is a set of ASEs and the constraints that bound their functions, to be used for a particular instance of communication. ACs are predefined packages that can be registered with ITU-T.

TCAP messages are specified in an application-independent manner [4–8,10]: the structure (syntax) of components is specified as a template, and the definition of individual operations (their parameters and detailed format) is left to the TC-user.

TCAP supports four classes of operations:

- Class 1: Success and failure are reported back
- Class 2: Only success is reported back
- Class 3: Only failure is reported back
- Class 4: No report

16.1.2 TCAP Protocol Stack

As shown in Fig. 16.1-1, TC is comprised of two sublayers:

1. Component sublayer, also called *TR-user*
2. Transaction sublayer (TR)

The component sublayer encapsulates components into *application protocol data units* (APDUs). The transaction sublayer handles the messages containing components, which are passed to/from the remote destination using the services of SCCP.

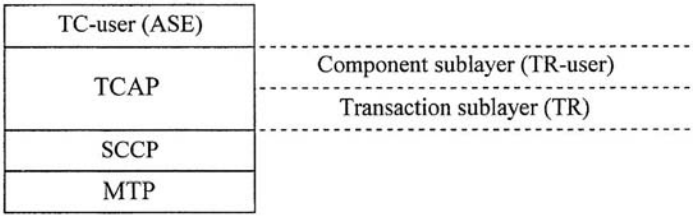


Figure 16.1-1. TCAP protocol stack.

Peer-level communication between TC-users (ASEs) located in remote nodes takes the form of *dialogs*, which consist of an exchange of components with the same Dialog ID. TCAP supports two kinds of dialogs:

1. Unstructured dialog
2. Structured dialog

Unstructured dialog allows the sending of one or more related components—in a UNIDIRECTIONAL message—to the remote end without expecting a response. It is used to implement class 4 operations.

Structured dialog is a true dialog with a beginning, a continuation, and an end, involving multiple messages. It allows the transfer of information between TC-users in full-duplex mode and is used to carry out class 1 through 3 operations. This kind of dialog is managed via the Dialog ID in TC dialog-handling primitives. The Dialog ID maps into the Transaction ID in TR primitives (see below).

Peer-level exchanges between component sublayers are called *transactions*. Transactions are carried out by the transaction sublayer as a service to the TR-user (i.e., the Component sublayer). Dialogs map one-to-one into transactions.

Communication within a node between the TC-user, TC sublayers, SCCP, and MTP is done via primitives, namely, TC primitives, TR primitives, N-unitdata primitives, and MTP primitives (Fig. 16.1-2) [4–8].

Figure 16.1-2 shows the SS7 entities involved in a transaction between ASEs belonging to two AEs located in node A and node B, respectively. The physical message path traverses the TCAPs, SCCPs, and MTPs at the nodes, and the signaling network, which transfers MTP's message signal units (MSUs) between the nodes.

16.1.3 Primitives

There are two types of TC primitives (see Table 16.1-1):

1. Dialog-handling primitives
2. Component-handling primitives

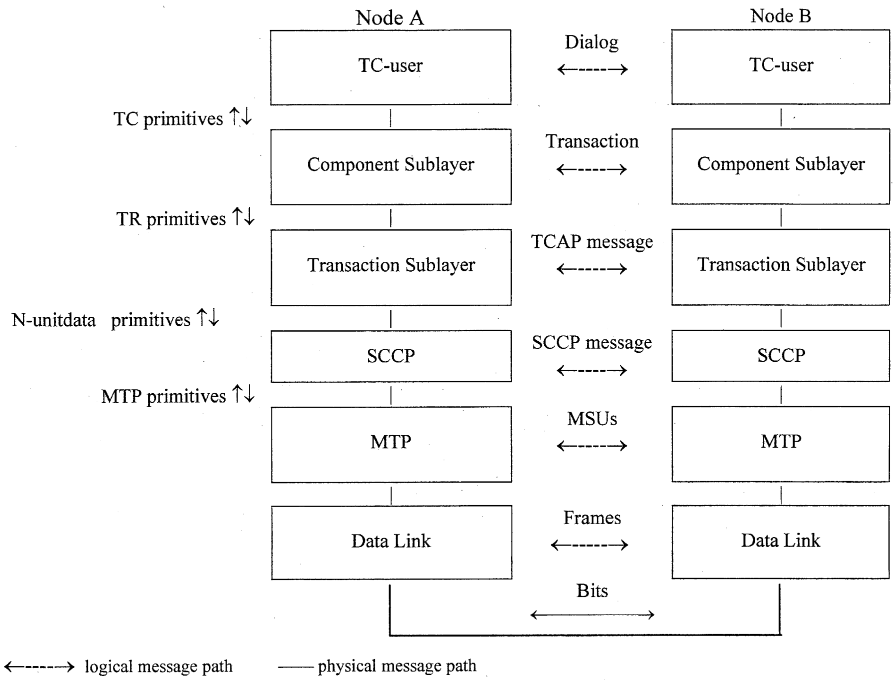


Figure 16.1-2. TCAP primitives, transactions, messages and message paths.

Dialog-handling TC primitives are used by the TC-user to trigger the sending of messages by the TR sublayer. These primitives map one-to-one into, and have the same names of, TR primitives as can be seen from Tables 16.1-1 and 16.1-2.

Component-handling TC primitives send individual components to the component sublayer.

Primitives can flow in either direction. When sent downstream from the TC-user or the TR-user, they are called *request* primitives and correspond to a locally generated request for an operation. Primitives sent upstream to the TR-user or to the TC-user are called *indication* primitives and correspond to the receipt of a remotely originated request for an operation or of a remotely originated response to an operation request.

The procedure that results in the exchange of TCAP messages containing operation-carrying components works as shown in Fig. 16.1-3. Individual components are sent from the TC-user to the component sublayer via component-handling TC primitives. The Component sublayer links components that are received with the same Dialog ID (assigned by the TC-user), as part of a dialog. When a dialog-handling TC primitive is received from the TC-user, the component sublayer sends all the linked components to the TR sublayer in the corresponding TR primitive. The TR sublayer then starts the transaction by sending a TCAP message containing the linked components to the remote end, where the process is reversed.

TABLE 16.1-1 TC Primitives

Type	Primitive	Function
Dialog handling	TC-UNI	Request indicates an unstructured dialog (UNI = unidirectional)
	TC-BEGIN	Begins a dialog
	TC-CONTINUE	Continues a dialog
	TC-END	Ends a dialog
	TC-U-ABORT	User terminates the dialog abruptly
	TC-P-ABORT	Dialog has been terminated by the service provider
	TC-NOTICE	Service provider has been unable to provide the requested services
Component handling	TC-INVOKE	Request an Operation to be executed
	TC-RESULT-L	[Last] Final or only segment result of an Operation
	TC-RESULT-NL	[Non-Last] Nonfinal part of segment result for an Operation
	TC-U-ERROR	Execution of the Operation has failed
	TC-L-CANCEL	[Local] Cancel Operation due to a timeout
	TC-U-CANCEL	Cancel Operation due to a TC-user decision
	TC-L-REJECT	[Local reject] Invalid component was received
	TC-R-REJECT	[Remote reject] Rejected by the remote Component sublayer
	TC-U-REJECT	Rejection by the TC-user
	TC-TIMER-RESET	Local TC-user wants to refresh a timer of an Operation invocation

Addressing. The addresses of the two AEs involved in a transaction, needed to route the TCAP message to the appropriate destination, follow the SCCP addressing scheme: point code and subsystem number (Section 14.2.4). Addresses are not carried in TCAP messages because message routing is handled by SCCP. The addresses are contained in two parameters, the *originating address* and the *terminating address*, which are sent from the AE to TC in a TC primitive, and from TC (unchanged) to SCCP in a TR primitive.

TABLE 16.1-2 TR Primitives

Primitive	Function
TR-UNI	Unstructured dialog (UNI = unidirectional)
TR-BEGIN	Begin transaction (structured dialog)
TR-CONTINUE	Confirm transaction or exchange user data (structured dialog)
TR-END	End transaction (structured dialog)
TR-U-ABORT	Abort transaction by user (structured dialog)
TR-P-ABORT	Abort transaction by service provider (structured dialog)
TR-NOTICE	Service provider unable to provide service (structured dialog)

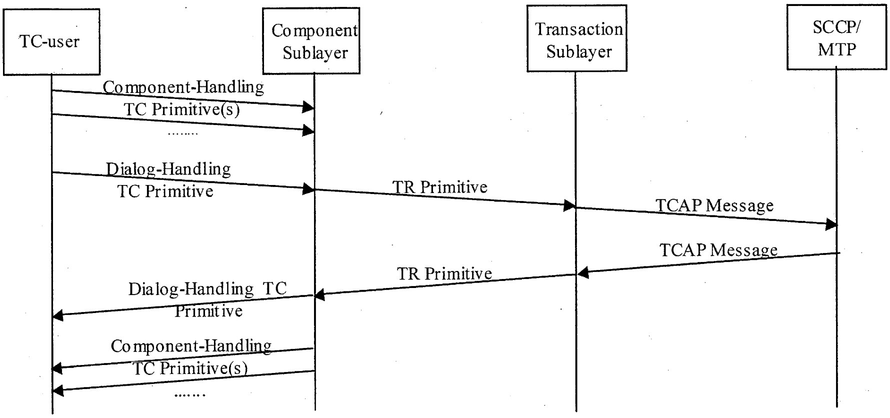


Figure 16.1-3. TCAP procedure for transferring components.

16.1.4 TCAP Messages

The TR sublayer can send four types of messages, shown in Table 16.1-3 with the TR primitives that cause them to be sent.

16.1.5 Components

Components are application protocol data units (APDUs) that carry operation requests and responses. Components are sent to the Component sublayer in component-handling TC primitives, from there to the TR sublayer in TR primitives, and from the TR sublayer to SCCP in N-unitdata primitives as the component portion of TCAP messages (Section 16.2.2).

There are four types of components, three of which are taken from the ROSE protocol and one that is specified for TCAP only (Table 16.1-4). The type of component sent to the TR sublayer is determined by the TC-user through the choice of component-handling TC primitive.

The RR-NL component (not in ROSE) allows transfer of data units longer than what can fit in a component. The maximum length of a MSU is 272 octets but, taking into account TCAP, SCCP, and MTP headers, the maximum length of a

TABLE 16.1-3 TCAP Messages

TCAP Message Type	Function	TR Primitive
UNIDIRECTIONAL	Unstructured dialog	TR-UNI
BEGIN	Start structured dialog	TR-BEGIN
CONTINUE	Confirm/continue structured dialog	TR-CONTINUE
END	End structured dialog	TR-END
ABORT	Abort structured dialog	TR-U-ABORT (user), TR-P-ABORT (service provider)

TABLE 16.1-4 Component Types

Component	Acronym	Function	ROSE	Component-Handling TC Primitive
Invoke	INV	Request an Operation	Y	TC-INVOKE
Return Result, Last	RR-L	Final response to an Invoke	Y	TC-RESULT-L
Return Error	RE	Error was detected	Y	TC-U-ERROR
Reject	RJ	Invalid Operation was received	Y	TC-U-REJECT TC-R-REJECT TC-L-REJECT
Return Result, Not Last	RR-NL	Nonfinal response to an Invoke	N	TC-RESULT-NL

component is about 240 octets. Most operations yield information that fits within one component and, if the information is longer, it is segmented and carried in one or more RR-NL components plus one RR-L component.

A TCAP message can carry multiple components.

16.1.6 Optional Dialog-Control APDUs

In addition to exchanging components, an AE can engage in an exchange of information with its remote peer to negotiate and agree upon the application context (Section 16.1.1). That takes place when a dialog-handling TC primitive includes a dialog control APDU containing an AC name as a parameter. The APDU with the AC name is passed to the remote node in the optional dialog control portion of a TCAP message (Section 16.2.2). The remote node can accept or reject the proposed AC name. Dialog control APDUs and their mapping to dialog-handling primitives are shown in Table 16.1-5.

When there is no AC negotiation, dialog control APDUs are not used and dialogs are managed by the transaction portion of TCAP messages (Section 16.2.2).

16.1.7 TCAP Message Sequence

Figure 16.1-4 shows examples of TCAP message sequences in transactions that involve ASE-1 and ASE-2. The transactions are initiated by ASE-1, which sends a Begin message with Invoke 1. The vast majority of transactions require just two messages: Begin and End (cases (a) and (b)). In case (a), the invoked operation results in an instruction, which is returned in an End message with Invoke 2. In

TABLE 16.1-5 Dialog Control APDUs

APDU	Acronym	Dialog-Handling TC Primitive
Dialog UNI [unstructured dialog]	AUDT	TC-UNI
Dialog Request	AARQ	TC-BEGIN
Dialog Response [accept]	AARE [accept]	TC-CONTINUE, TC-END
Dialog Response [rejected]	AARE [rejected]	TC-U-ABORT
Dialog Abort Dialog	ABRT	TC-U-ABORT

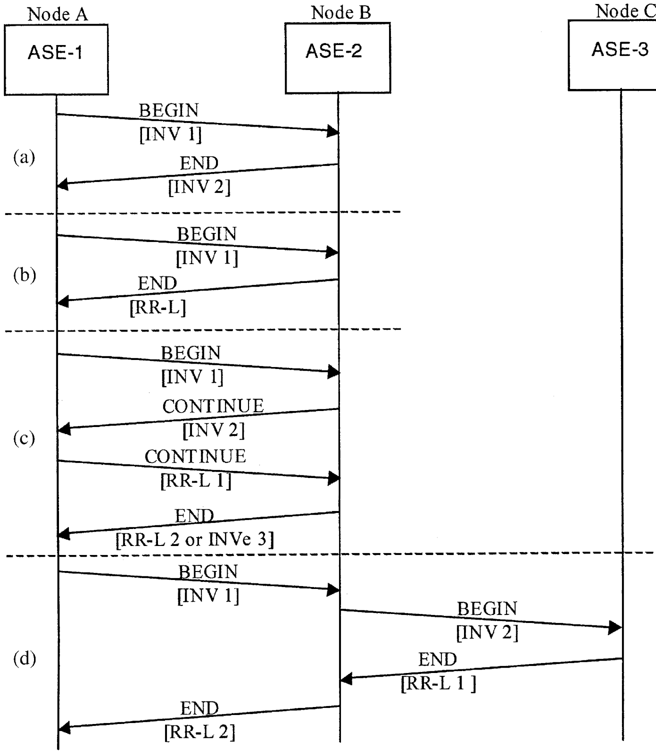


Figure 16.1-4. Transaction examples.

case (b), the invoked operation yields information, which is returned in an End message with a RR-L component.

In case (c), ASE-2 needs more information from ASE-1 before it can execute Invoke 1. It therefore returns a Continue message, whose Invoke 2 requests the information. After executing Invoke 2, ASE-1 sends a Continue message, in which RR-L 1 contains the requested information. ASE-2 now executes Invoke 1 and returns a RR-L or an Invoke in an End message.

Case (d) shows a chain of transactions, involving three ASEs. ASE-2 cannot execute Invoke 1 without information from ASE-3. It therefore initiates a transaction with ASE-3, sending Invoke 2. ASE-3 executes Invoke 2, obtains a result, and includes it in the RR-L 1 of its End message. ASE-2 then executes Invoke 1 and sends ASE-1 an End message with the result in RR-L 2.

16.2 TCAP FORMATS AND CODING

The detailed format for TCAP messages is specified in ITU-T recommendations using ASN.1, a notation similar to a computer language, which produces definitions that can be processed by software [14–17]. That has resulted in the availability of

ASN.1 compilers that can greatly speed up the creation of complex, error-free specifications. To take advantage of the flexibility offered by computer-based tools, ASN.1 provides complete separation of the bit map definition ("bits on the line") from the syntax (higher level definition of fields) used by a protocol. The bit-independent definition is called *Abstract Syntax*, and the translation into bits is called *encoding*. Several types of encoding algorithms are available and TCAP uses the most straightforward, called *basic encoding rules* (BER) [16].

Because encoding is now rarely done by hand, the actual bit map for TCAP messages is not described in this chapter. We shall describe only the basic structure of messages, as a sequence of information elements (IEs) [6].

16.2.1 Information Elements (IEs)

Each information element (or *abstract field*, in ASN.1 terminology) in a TCAP message is encoded, at the bit level, as three fields (see Fig. 16.2-1):

1. Tag
2. Length
3. Contents

The Tag field indicates the IE type, the Length field indicates the number of octets in the Contents field, and the Contents field contains the information to be transferred. Tag and Length are generated by the encoding algorithm and are not specified by the Abstract Syntax.

A TCAP IE can be of two types:

1. Primitive IE
2. Constructor IE

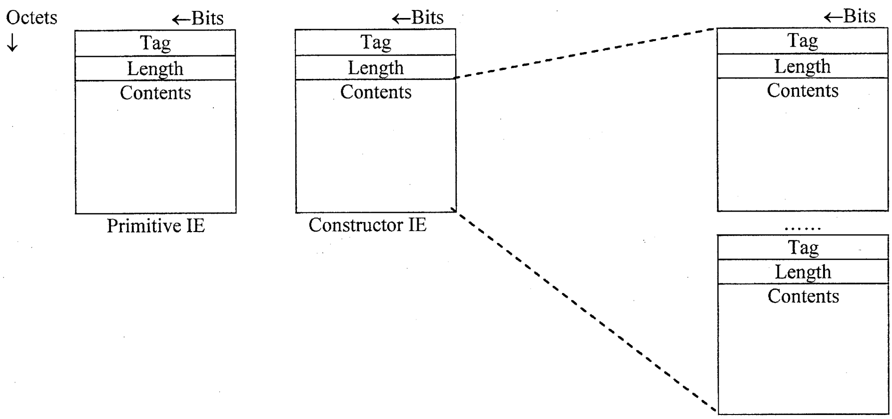


Figure 16.2-1. Structure of TCAP information elements. (From Rec. Q.773. Courtesy of ITU-T.)

In primitive IEs, the Contents field contains only one value—the information carried by that IE. Primitive IEs should not be confused with the primitives that pass information between layers/sublayers in the same node (Fig. 16.1-2).

In constructor IEs, the Contents field contains multiple IEs, which may be primitive or constructor (nesting), as shown in Fig. 16.2-1.

The primitive and constructor IE concepts are carried over from ITU-T Recommendation X.209 (Data Communication Networks).

16.2.2 TCAP Message Format

A TCAP message is a constructor information element containing a number of mandatory (M) and/or optional (O) IEs [6], in a nested arrangement. We shall see that the term *parameter* is used also but has a restricted meaning.

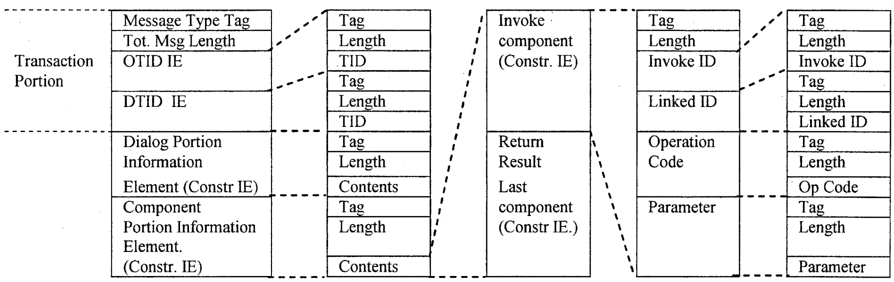
A TCAP message consists of three parts (see Fig. 16.2-2):

1. Transaction Portion
2. Dialog Control Portion
3. Component Portion

Transaction Portion. Like all constructor IEs, a TCAP message starts with a Type Tag that identifies the message type and a Length field that covers the total length of the message. Those two fields form an overall "wrapper" for the message and are part of the Transaction Portion. The Transaction Portion may contain other primitive IEs (depending on the type of message), such as the Transaction ID that is used to relate messages that are part of the same transaction (dialog).

Dialog Control Portion. This is an optional constructor IE, which, if present, contains a dialog control APDU (Section 16.1.6).

Component Portion. This is a constructor IE that contains one or more components. Individual components are also constructor IEs containing multiple primitive IEs.



Constr.: Constructor Tot. Msg.: Total Message OTID: Origination Transaction ID DTID: Destination Transaction ID

Figure 16.2-2. Structure of a TCAP message.

TABLE 16.2-1 Fields in TCAP Messages

Message Portion	Field	Type	Function
Transaction	Message Type Tag and Length	C	Identifies the message type and length
	Origination Transaction ID	P	Manages dialogs
	Destination Transaction ID	P	Manages dialogs
	P-Abort cause	P	Describes reason for abort by TR sublayer
Dialog	Dialog Portion Tag and Length	C	Delimits the Dialog portion
	Dialog PDU	C	Sets up the Application Context[a]
Component	Component Portion Tag and Length	C	Delimits the Component Portion
	Component Type Tag and Length	C	Identifies a particular component
	Invoke ID	P	Identifies a particular invocation of an Operation
	Linked ID	P	Response to an invocation with another invocation
	Operation Code	P	Identifies the Operation
	Parameter	P/C	From the Operation (one or more)
	Error Code	P	From the Operation
	Problem Code[b]	P	Reason for rejecting the component
	General Problem		Received component cannot be processed
	Invoke Problem		Received Invoke cannot be processed
	Return Result problem		Received RR cannot be processed
	Return Error problem		Received RE cannot be processed

[a]Details of this constructor IE can be found in [6].
[b]Four categories of problems based on Tag values.

Note: C, constructor information element; P, primitive information element (their Tag and Length fields are not shown).

TCAP message IEs are shown in Tables 16.2-1, 16.2-2, and 16.2-3. Figure 16.2-2 illustrates the structure of a TCAP message with an example (Continue) containing two components. In the figure, only the Invoke component is expanded into its primitive IEs; the primitive IEs for the RR-L component can be found in Table 16.2-3.

TABLE 16.2-2 Information Elements in a TCAP Message

	IEs↓ → Message Type	Unidirectional	Begin	End	Continue	Abort
Transaction Portion IEs	Origination Transaction ID		M		M	
	Destination Transaction ID			M	M	M
	P-Abort cause					O
	Dialog Portion	O	O	O	O	O
	Component Portion	M	O		O	

Note: M, mandatory; O, optional.

TABLE 16.2-3 Information Elements in Components

IEs↓ → Component	INV	RR-L	RE	RJ	RR-NL
Invoke ID	M	M	M	M	M
Linked ID	O				
Operation Code	M	O			O
Parameter	O	O	O		O
Error Code			M		
Problem Code				M	

Note: O, optional; M, mandatory.

Operations. The main purpose of TCAP messages is to exchange operation requests and responses between ASEs that are part of distributed applications such as INAP.

Operations in TCAP are used as a template (Chapter 14) [6]. The generic template is borrowed by TCAP from the ROSE protocol, which originally defined it is as a *macro* and more recently as an *information object class* [11] (in both cases with the ASN.1 notation). The TCAP recommendations still use the macro notation [6] but, either way, the format is the same. The names of the different types of operations, errors, and their parameters are left for the TC-user application to define and assign values to.

Operation parameters are exchanged in TCAP components, with different components containing different parameters:

- All components contain the Operation code.
- Invoke contains the ARGUMENT and the LINKED parameters.
- Return Result contains the RESULT parameters.
- Return Error contains the ERRORS parameters.
- Reject does not send any operation parameter (except for the Op Code). A Reject means that the request to perform the operation is rejected, so there are no results or even errors to report. Parameters sent with Reject are TCAP error parameters.

16.2.3 Coding of Tag and Contents Fields

Tag Field. An IE Tag field consists of one or more octets and has three fields (Fig. 16.2-3).

Class Field. Bits H, G of octet 1 indicate the IE class:

Class	H	G
Universal	0	0
Application-wide	0	1
Context-specific	1	0
Private use	1	1

Octets ↓	H	G	F	E	D	C	B	A ←Bits
1	Class		Type	Tag Code				

	H	G	F	E	D	C	B	A
1	Class		Type	1	1	1	1	1
2	0	Tag Code						

	H	G	F	E	D	C	B	A
1	Class		Type	1	1	1	1	1
2	1	Tag Code						
.	.....							
.	1	Tag Code						
n	0	Tag Code						

Figure 16.2-3. Tag fields. (From Rec. Q.773. Courtesy of ITU-T.)

The class points to where the IE is specified. Universal IEs are specified in ITU-T Rec. X.409. IEs in this class are used primarily in data communication networks and some are also used in TCAP. Application-wide IEs are specified by ITU-T in [6]. They are used in all TCAP applications defined by ITU-T. Context-specific IEs are specified within the context of the next higher constructor IE. A context-specific Tag value thus has different meanings, which depend on the class of the constructor. Private use IEs are defined by various national organizations.

The "class" concept enables the various standards organizations to define IEs independently of each other.

Type Field. Bit F of octet 1 indicates whether the IE is constructor (F = 1) or primitive (F = 0).

Tag Code Field. The Tag code identifies the IE. Tag codes that range from 00000 through 11110 (decimal 0 through 30) are coded as one-octet Tags, in bits E to A. Tag codes higher than 31 are coded as multi-octet Tags: bits E to A in octet 1 are set to 11111, and bits G to A in the following octets hold the Tag code. Bit H of the second and higher octets is set to 0 if that octet is the last one; otherwise it is set to 1.

TABLE 16.2-4 Examples of Tag Codes

IE	Type ↓ → Tag Bits	H	G	F	E	D	C	B	A
TCAP Message	UNIDIRECTIONAL	0	1	1	0	0	0	0	1
	BEGIN	0	1	1	0	0	0	1	0
	CONTINUE	0	1	1	0	0	1	0	1
	END	0	1	1	0	0	1	0	0
	ABORT	0	1	1	0	0	1	1	1
Problem Code	General problem	1	0	0	0	0	0	0	0
	Invoke problem	1	0	0	0	0	0	0	1
	Return Result problem	1	0	0	0	0	0	1	0
	Return Error problem	1	0	0	0	0	0	1	1

Source: Rec. Q.773. Courtesy of ITU-T.

TABLE 16.2-5 Examples of Coding of the Contents Fields

IE	Function	Bits H thru A
P-Abort Cause	Unrecognized message type	0000 0000
	Unrecognized Transaction-ID	0000 0001
General problem	Unrecognized component type	0000 0000
	Mistyped component (the component is not properly structured)	0000 0001
Invoke problem	Duplicate Invoke-ID	0000 0000
	Unrecognized Operation code	0000 0001
	Mistyped parameter (Tag code not recognized)	0000 0010
Return-Result problem	Unrecognized Invoke-ID	0000 0000
	Mistyped parameter	0000 0010
Return-Error problem	Unrecognized Invoke-ID	0000 0000
	Unrecognized error	0000 0010

Table 16.2-4 lists some examples of Tag codes of the IEs discussed in this section. The Tags of IEs in the Transaction Portion are coded “application-wide.” The Tags of IEs in the Component Portion are coded “context-specific.” A few Tags are coded as “universal.”

Contents Field. Examples of IE Contents field coding for primitive IEs are shown in Table 16.2-5. More detail can be found in [6].

16.3 TRANSACTION AND INVOKE IDENTITIES

At any point in time, a number of transactions can be active at a node. Moreover, an AE can be involved in several concurrent transactions, and a transaction can consist of more than one remote operation.

The transfer of transaction messages has been described in Section 15.3.4. We now discuss the transaction and component IDs, which are used to keep track of individual transactions and operations.

16.3.1 Transaction Identifiers

The TC-user (ASE) at a node uses a local Dialog ID parameter to identify a dialog with its peer at the remote node. The TC entity associated with the TC-user maps the Dialog ID to a Transaction Identifier (TID), a reference that identifies a transaction at the TC level. TIDs are assigned independently, from a pool, by the TCs at either end and exchanged in TCAP messages. Figure 16.3-1 shows an example of how the TID is used. When ASE-1 at node A initiates a transaction with ASE-2 at node B, TC-A assigns TID = p from its pool to the transaction and associates it to the Dialog ID from ASE-1.

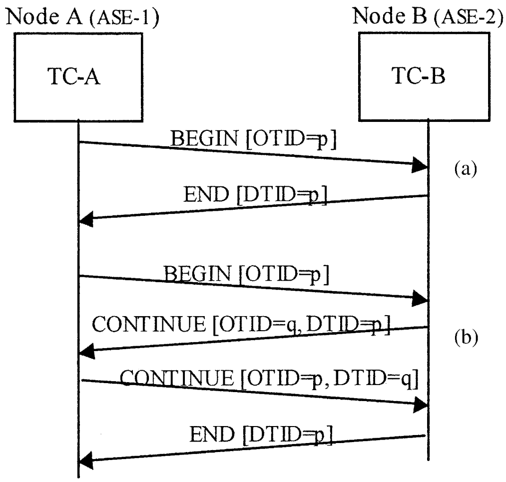


Figure 16.3-1. Transaction identifiers.

The Begin message from TC-A includes an *Originating TID* field (OTID) with the value OTID = p. TC-B allocates TID = q to the new transaction at its node and associates it and the remote TID = p to a Dialog ID with ASE-2. TC-B then passes the component(s) in the Begin message to ASE-2. The transaction is now identified for TC-A by TID = p, and for TC-B by TID = q.

In case (a), ASE-2 requests TC-B to send an End message, which terminates the transaction. TC-B then returns TID = q to its TID pool. Its End message includes the field *Destination TID* (DTID) with the value of "p." This identifies the transaction for TC-A, which passes the component(s) in the message to ASE-1 (the ASE that is currently associated with TID = p). Since the transaction has ended, TC-A then returns TID = p to its TID pool.

In case (b), ASE-2 continues the transaction. The Continue message from TC-B includes OTID = q and DTID = p. TCAP-A then adds remote TID = q to the record that associates the ASE-1 Dialog ID with the local TID = p. In this example, ASE-1 also continues the transaction, and the Continue message from TC-A includes OTID = p and DTID = q.

The procedures for the End message of TC-B are as described above.

16.3.2 Component Identifiers

A transaction can involve several Invoke components, each identified by an *Invoke ID* (IID), to allow for multiple invocations of the same operation. IID is assigned by the TC-user from a pool and identifies a particular invocation of an operation to the TC-users at both ends.

Four examples are shown in Fig. 16.3-2. In example (a), the allocated IID = x is included in the Invoke component of the Begin message from TC-A. The same IID value is included in the Return-Result of the End message sent by TC-B. In general,

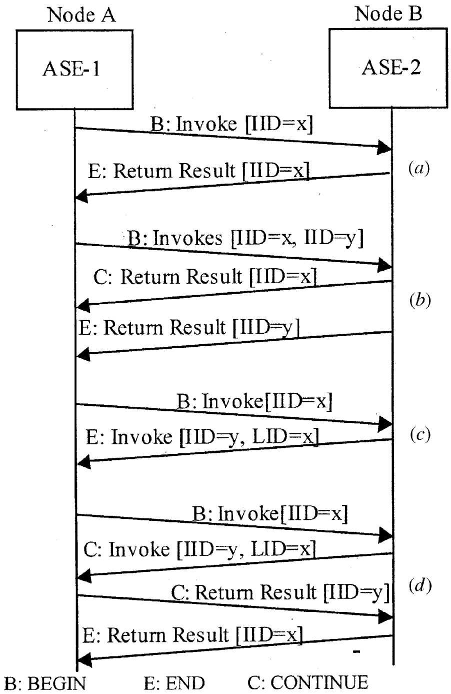


Figure 16.3-2. Invoke identifiers.

a Return-Result, Error, or Reject component includes an IID that "reflects" the IID of the received Invoke.

In example (b), the Begin message contains two Invokes (IID = x and IID = y). ASE-2 executes the first operation and sends a Return-Result component with IID = x. It then executes the second operation, and the Return-Result component includes IID = y.

In example (c), the Invoke of ASE-1 is a request for instructions and includes IID = x. ASE-2 responds with an Invoke of its own, which specifies the requested instructions. The Invoke includes IID = y, assigned by ASE-2 to its Invoke, and *Linked Identifier* LID = x, which refers to the Invoke of ASE-1.

Finally, in example (d), ASE-2 responds to the Invoke of ASE-1 with an Invoke that requests more information. In the Continue message from TC-B, IID = y and LID = x refer to the Invokes of ASE-2 and ASE-1. In the Return-Result component from ASE-1, IID = y refers to the Invoke of ASE-2, and the Return-Result from ASE-2 includes IID = x.

16.4 U.S. NATIONAL TCAP

The version of TCAP defined by the American National Standards Institute (ANSI) for use in fixed networks in the United States is very similar to the TCAP defined by ITU-T. The differences, which are relatively small, are outlined below.

TABLE 16.4-1 ITU-T IEs and Equivalent ANSI DEs

	Information Element	C/P	Data Element	C/P
Transaction Portion	Unidirectional	C	Unidirectional	C
	Begin	C	Query	C
	Continue	C	Conversation	C
	End	C	Response	C
	Abort	C	Abort	C
	Origination Transaction ID	P	TID	P
	Destination Transaction ID	P	TID	P
	P-Abort cause	P	P-Abort cause	P
	Component Portion	C	Component Sequence	C
Component Portion	Invoke	C	Invoke	C
	Return Result, Last	C	Return Result, Last	C
	Return Result, Not Last	C	Return Result, Not Last	C
	Return Error	C	Return Error	C
	Reject	C	Reject	C
	Invoke ID	P	Component ID	P
	Linked ID	P	Component ID	P
	Operation Code	P	Operation Code	P
	Parameter	P/C	Parameter	P/C
	Error Code	P	Error Code	P
	Problem Code	P	Problem Code	P

Note: C, constructor; P, primitive.

16.4.1 Terminology

In ANSI documents, *Data Elements* (DEs), *Package*, and *Identifier* replace *Information Element*, *Message*, and *Tag*. Table 16.4-1 lists a number of ITU-T TCAP terms for IEs and their ANSI equivalent, or near equivalent.

16.4.2 Transaction and Component IDs

ANSI does not have separate DEs for originating and destination transaction IDs. In query and response packages, the transaction ID (TID) value field holds one 32-bit integer, which represents the originating or destination TID value. In conversation packages, the TID field holds two 32-bit integers. The first integer represents the originating TID, and the second represents the destination TID.

ANSI uses the *Component-ID* data element in lieu of the Invoke-ID and Linked-ID information elements. If an ASE sends an Invoke in response to a received Invoke, the sent Invoke includes two Component-IDs representing, respectively, the Invoke-ID and the Linked-ID.

16.4.3 Identifier Coding

The U.S. TCAP uses *context-specific* (H, G = 1, 0) and *private-use* (H, G = 1, 1) identifier (Tag) classes of ITU-T (Section 16.2.3), However, the private-use class is redefined as *national TCAP/private TCAP.*

National TCAP refers to TCAP applications in ‘fixed” telecommunication networks, which are specified by ANSI and Telcordia.

Private TCAP applications are defined by the Electronic Industries Alliance (EIA) and the Telecommunications Industry Association (TIA) for use in mobile networks.

For identifiers (Tags) with H,G = 1,1 in octet 1 (see Fig. 16.2-3), the values in the identifier code field are divided into separate ranges, for national and private use. All values in one-octet identifier fields are assigned to national TCAP. In multiple-octet fields, the value of bit G in octet 2 indicates whether the code is assigned to national TCAP (G = 0) or private TCAP (G = 1).

In general, if an ITU-T information element is coded as “application-wide,” the corresponding ANSI data element is coded as “national TCAP.”

Further details on Identifier coding of national TCAP DEs can be found in [9].

16.4.4 Value Coding of Data Elements

National TCAP Error-Code DE. ANSI has specified a one-octet value field for the Error-Code DE in Return-Error packages. The value indicates the reason for sending the Return-Error:

Unexpected component sequence	0000 0001
Unexpected data value	0000 0010
Unavailable network resource	0000 0011
Reply overdue	0000 0101
Data not available	0000 0110

National TCAP Problem DE. The Problem-Code DE has a two-octet value field. The first octet (known as *problem code*) indicates the nature of the problem (general problem, invoke problem, etc.) The second octet (*problem specifier*) indicates a specific problem (unrecognized component duplicate Invoke ID, incorrect parameter, etc.).

Further details on the coding of value fields for national TCAP data elements can be found in [9].

National TCAP Constructor and Primitive DEs. Table 16.4-1 shows that if an IE is a constructor (C), the corresponding DE is also a constructor, and the same holds for primitive (P) IEs and DEs. Some minor exceptions (not shown in the table) exist and further details can be found in [9].

16.5 ETSI TCAP

The European Telecommunications Standard Institute (ETSI) has adopted the ITU-T recommendations virtually unchanged. Minor differences are spelled out in [10].

16.6 ACRONYMS

AC	Application context
AE	Application entity
ANSI	American National Standards Institute
APDU	Application protocol data unit
ASE	Application service element
ASN.1	Abstract Syntax Notation One
C	Constructor information element
CCITT	International Telegraph and Telephone Consultative Committee
DE	Data element
DTID	Destination TID
EIA	Electronic Industries Alliance
ETSI	European Telecommunications Standards Institute
ID	Identity
IE	Information element
IID	Invoke ID
INAP	Intelligent Network Application Part
ISUP	ISDN User Part
ITU	International Telecommunication Union
ITU-T	Telecommunication Standardization Sector of ITU
LID	Linked identifier
M	Mandatory
MAP	Mobile Application Part
MSU	Message signal unit
MTP	Message Transfer Part
NA	Not applicable
O	Optional
OTID	Originating TID
P	Primitive information element
ROSE	Remote Operations Service Element
ROS	Remote operations service
RR-L	Return result last
RR-NL	Return result not last
SCCP	Signaling Connection Control Part
SS7	Signaling System No. 7
TC	Transaction capabilities
TC-U-. . .	TC-user-. . .
TC-L-. . .	TC-local-. . .
TC-R-. . .	TC-remote-. . .

TCAP	Transaction Capabilities Application Part
TIA	Telecommunications Industry Association
TID	Transaction identifier
TR	Transaction
TR-U-. . .	TR-user-. . .
TR-P-. . .	TR-provider-. . .
UNI	Unidirectional

16.7 REFERENCES

1. A. R. Modaressi, and R. A. Skoog, Signaling System No. 7: A Tutorial, *IEEE Commun. Mag.*, **28**, No. 7, July 1990.
2. R. Mantefield, *Common Channel Signalling*, Peter Peregrinus Ltd., London, 1991.
3. T. W. Johnson, B. Law, and P. Anius, CCITTT Signalling System No. 7: Transaction Capabilities, *Br. Telecommun. Eng.*, **7**, Apr. 1988.
4. *Functional Description of Transaction Capabilities*, Rec. Q.771, ITU-T, Geneva, 1997.
5. *Transaction Capabilities Information Element Definition*, Rec. Q.772, ITU-T, Geneva, 1997.
6. *Transaction Capabilities Formats and Encoding*, Rec. Q.773, ITU-T, Geneva, 1997.
7. *Transaction Capabilities Procedures*, Rec. Q.774, ITU-T, Geneva, 1997.
8. *Guidelines for Using Transaction Capabilities*, Rec. Q.775, ITU-T, Geneva, 1997.
9. *American National Standard for Telecommunications-Signaling System No. 7 (SS7)—Transaction Capabilities*, ANSI T.114, ANSI, New York, 1992.
10. *Integrated Services Digital Network (ISDN); CCITT Signalling System No. 7 Transaction Capabilities Application Part (TCAP) Version 2*, ETS 300 287, ETSI, Sophia Antipolis, France, 1993.
11. *Remote Operations: Concept, Model and Notation*, Rec. X.880, ITU-T, Geneva, 1994.
12. *Remote Operations: Model, Notation and Service Definition*, Rec. X.219, ITU-T, Geneva, 1988.
13. *Remote Operations: Protocol Specification*, Rec. X.229, ITU-T, Geneva, Nov. 1988.
14. *Information Technology—Abstract Syntax Notation One (ASN.1): Specification of Basic Notation*, Rec. X.680, ITU-T, Geneva, 2002.
15. *Information Technology—Abstract Syntax Notation One (ASN.1): Information Object Specification*, Rec. X.681, ITU-T, Geneva, 2002.
16. *Information Technology—ASN.1 Encoding Rules: Specification of Basic Encoding Rules (BER), Canonical Encoding Rules (CER) and Distinguished Encoding Rules (DER)*, Rec. X.690, ITU-T, Geneva, 2002.
17. *Information Technology—ASN.1 Encoding Rules: Specification of Encoding Control Notation (ECN)*, Rec. X.692, ITU-T, Geneva, 2002.
18. *Remote Operations: OSI Realizations—Remote Operations Service Element (ROSE) Service Definition*, Rec. X.881, ITU-T, Geneva, 1994.
19. *Remote Operations: OSI Realizations—Remote Operations Service Element (ROSE) Protocol Specification*, Rec. X.882, ITU-T, Geneva, 1994.

17

TRANSACTIONS IN INTELLIGENT NETWORKS

This chapter describes an important application of transactions and includes information on the operations and parameters for this application.

17.1 INTRODUCTION TO INTELLIGENT NETWORKS

17.1.1 History

The introduction of stored-program controlled exchanges in telecommunication networks has enabled the telecoms to offer call forwarding, call waiting, and other supplementary services to their customers [1]. These services are *exchange-based*: the (software) logic and data for the services reside in the exchanges.

Telecommunication networks are evolving into *intelligent networks* that offer, in addition to exchange-based services, a number of services whose logic and data reside in a small number of centralized information sources that can be queried (interrogated) by the exchanges.

The first steps toward intelligent networks (INs) were made by the Bell System in the early 1980s, for 800-call services [l]. 800 numbers contain no routing information and have to be translated into "routing" numbers that have the North American numbering plan (NANP) format AC(3)-*NXX-XXXX* (Section 1.2.1). The translation data are stored in network databases. This approach was taken because the translations require a large amount of data that has to be updated frequently. It is much simpler to add or change an entry in a few centralized databases than to enter these changes into every local exchange.

The queries by the exchanges, and the responses from the databases, were transferred in CCIS "direct signaling" messages (Chapter 5).

Signaling in Telecommunication Networks, Second Edition, by John G. van Bosse and Fabrizio U. Devetak

After the divestiture of the Bell system, the new regional telecoms started to deploy *service control points* (SCPs) with databases for 800 calling and *alternate billing services* [2]. Messages between exchanges and SCPs are transferred by the SS7 signaling network (Section 15.3.4).

The evolution to the present intelligent networks, known in the United States as *advanced intelligent networks* (AINs), started in the 1980s [3,4]. The logic and data for AIN services is stored in AIN SCPs, and exchanges obtain their call-handling instructions by executing transactions with these SCPs.

The definition of AIN is the result of the combined efforts of Bellcore (now Telcordia), the former Bell system regional telecoms, and equipment manufacturers. During the same time frame, ITU-T (International Telecommunication Union) started a phased standardization process for IN architecture and capabilities [5]. Each phase is known as an IN capability set (CS). The definition of CS-l, CS-2, CS-3, and CS-4, has been completed [6–8].

AIN Releases. AIN Release 1 outlines the general long-term objectives for AIN in the United States [9,10]. The detailed specifications have been published as a sequence of AIN releases (0.0, 0.1, and 0.2). Release 0.0 [11–13] has been used mainly for laboratory and service trials. Compared with its predecessor, AIN 0.1 is more closely aligned, in capabilities and terminology, with the international (ITU-T) standards. It is the first true step toward AIN 1. The specifications for AIN 0.1 have been published by BellSouth [14,15] and by Bellcore [16,17]. In this chapter, the focus is on AIN 0.1. AIN 0.2 is now available as well [18,19].

17.1.2 AIN 0.1 Architecture

The principal entities in an AIN 0.1 network are shown in Fig. 17.1-1.

Service Control Point. This is a data processor that stores the logic and data for AIN services. It responds to queries (requests for instructions) from SSP exchanges. The queries and responses are transferred by the SS7 signaling network.

Service Switching Point (SSP). This is an exchange that provides AIN services, by querying an SCP.

A SSP can be a local, tandem, or a combined local/tandem exchange. The present trend is to install SSP capabilities in all local exchanges, and this chapter limits the discussion to SSP *local* exchanges.

The SSP in Fig. 17.6-1 serves analog and digital (ISDN) subscribers. We shall refer to them as *customers* (C). SSP also serves private branch exchanges (PBX—see Section 1.1.2). The PBXs are attached to SSP by trunk groups that are known as *private-facility groups* (PFGs). Other PFGs are the *tie-trunk* groups, leased by the telecom to business customers who have several business locations. These trunk groups interconnect exchanges of the public network but are dedicated to calls made by lines of the individual business customers.

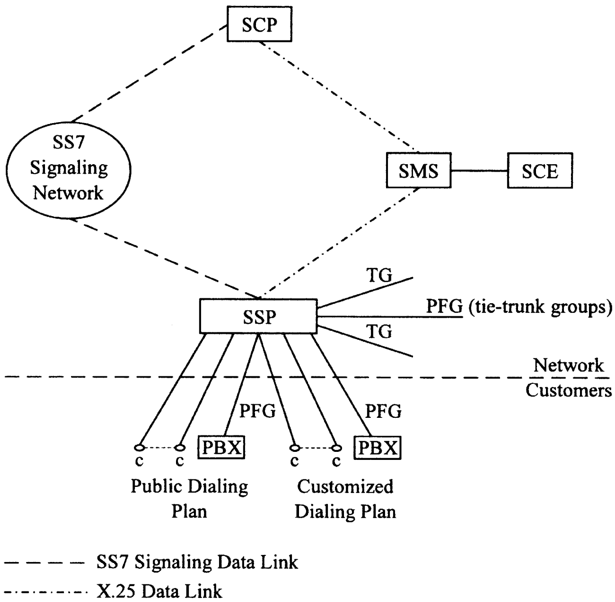


Figure 17.1-1. AIN 0.1 architecture.

Numbering Plans. Most customers and PBXs use the public North American numbering plan (NANP). However, at the request of a multilocation business customer, a telecom can provide a *customized* dialing plan that is "tailored" to the customer's requirements.

AIN provides services to customers using the NANP and to customers who have a customized dialing plan.

Service Management System (SMS). This system enables a telecom to update and maintain the service logic and data in its SCPs. The SMS validates inputs from the telecom and then updates the affected SCPs. SMS is also capable to test the AIN, and to collect AIN traffic data from SSPs. The SMS communicates with the SCPs and SSPs in its network by X.25 data links.

Service Creation Environment (SCE). This system interfaces with SMS and contains software tools that enable the telecoms to develop and test new AIN services.

17.1.3 AIN Service Switching Point

The conversion of an exchange into a SSP requires software and hardware additions.

SSP Software. Three main software functions have to be installed. In the first place, a SSP needs an AIN *application service element* (ASE) to communicate with a "peer" ASE in a SCP (Section 16.1.1).

In the second place, a SSP has to determine whether a call requires assistance from a SCP. This is done in a software procedure known as *triggering*.

In the third place, AIN transactions end with a response message from SCP that contains call-handling instructions. Software for the execution of these instructions has to be installed in the SSPs.

The triggers and the call-handling instructions are defined in a "generic" manner (not specific to a particular AIN service). A SSP does not know anything about the individual services: it merely informs SCP when it has encountered a trigger and executes the instructions received from SCP.

SSP Hardware. A SSP has to be equipped with a group of *intelligent network service circuits* (INSCs)—see Fig. 17.1-2. These circuits have two functions. In the first place, they play customized recorded announcements, whose contents are specified by SCP. In the second place, they can receive *dual-tone multifrequency* (DTMF) digits (Section 3.3.3). This enables a SCP to interact with the calling party. On instructions from SCP, the SSP processor connects an INSC to a calling line or incoming trunk and orders it to play a specified announcement, which prompts the calling party to respond with one or more digits. The SSP processor collects the received digits and then sends them to SCP.

17.1.4 AIN Service Control Point

The AIN SCP is a data processor that has the logic and data needed for AIN services. It includes one or more ASEs for communications with the SSPs in its network.

The advantages of SCPs to a telecom are twofold.

Telecom Programming. Before AIN, when new features needed to be added in a telecommunication network, the telecom had to purchase a new version of exchange

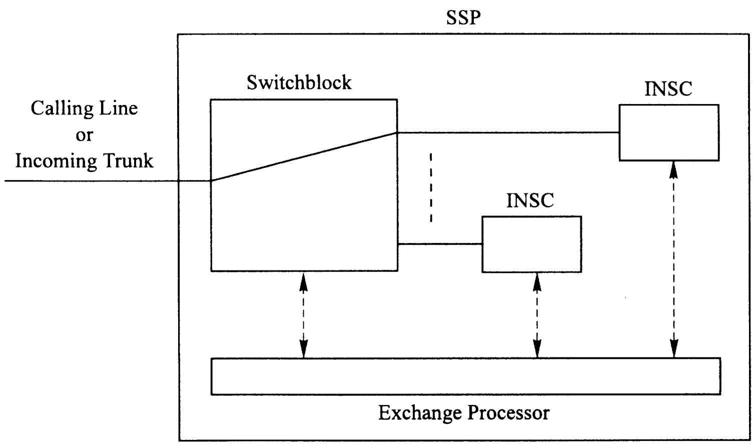


Figure 17.1-2. Intelligent network service circuits.

software from the equipment manufacturer and then install and test it in all exchanges.

AIN services can be programmed by telecom personnel, using the service creation environment. This gives the individual telecoms the freedom to design their own service features.

Also, the tests of a new AIN feature can be done with the SCP and just one of the SSPs in the network.

Flexibility. SCPs are designed to provide great flexibility in the definition of AIN services. A service can be designed to take into account the time-of-day, and day-of-week, of the call, and the locations of the calling and called parties.

Caller interaction is another important contributor to service flexibility. SCP can prompt the caller to send DTMF digits, which indicate a particular service option for the call, or an authorization code that is checked at SCP, to verify that the caller is entitled to receive the requested service.

17.1.5 AIN 0.1 Services

The services defined for AIN 0.1 involve transactions during the setup of connections only and apply to calls between two parties. Services requiring transactions during the conversation phase of a call and services for multiparty calls are supported in later AIN releases.

The emphasis is on capabilities to screen outgoing and incoming calls and to route and charge calls.

17.2 CALL MODELS AND TRIGGERS

When a SSP exchange receives a call that involves an AIN service, it initiates a transaction with SCP, to obtain call-handling instructions. The mechanism by which the SSP determines that it needs to query SCP is known as *triggering*. We say that a SSP *launches a query* (initiates a transaction with a SCP) when it "encounters a trigger" during the processing of a call. Triggers can be encountered by SSP at various points in the processing of a call. In order to explore the various trigger types, we need a call-processing model.

17.2.1 Basic Call Model

The call-processing models used in documents on intelligent networks are known as *basic call models* (BCMs) and consist of an originating and a terminating part [5,6,13].

Originating BCM (O-BCM). This chapter uses the simplified originating BCM shown in Fig. 17.2-1. The model consists of a number of call states, starting with the detection of an origination attempt by SSP (A). For simplicity, the model

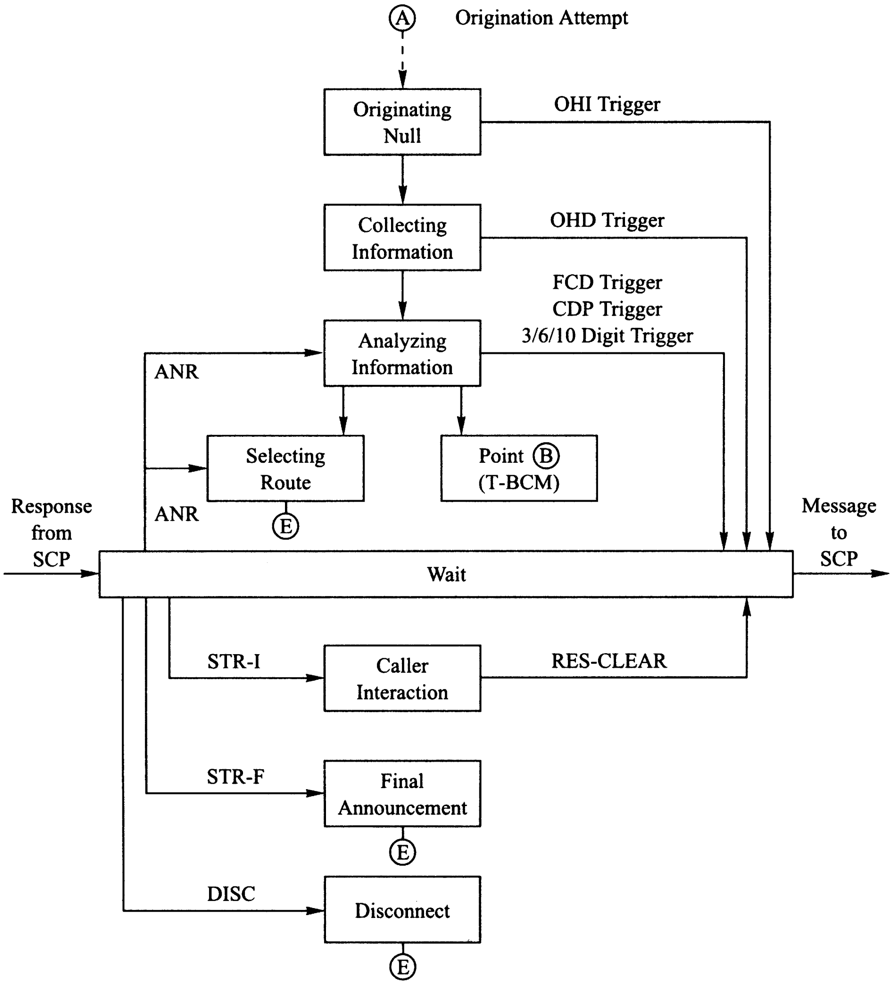


Figure 17.2-1. Originating BCM.

ends at points (E), which are the points where the call destination—either a line attached to the SSP or a route set to the destination exchange—has been determined.

We first describe "normal" call processing (no trigger encountered). Idle lines and trunks are supervised by the SSP. When an origination attempt is detected, say, an off-hook on an analog subscriber line (point A), the call process starts at the *originating null* state and moves to the *collecting information* state. In this state, SSP collects the information received from the calling customer (the digits dialed by an analog subscriber, or the digits in the SETUP message from an ISDN user).

The call then moves to the *analyzing information* state, where the call destination is determined, from the analysis of the digits of the called number. If the destination is "local" (the called customer is served by SSP), the call moves to point B in the terminating BCM (see below). If the destination is at another exchange, the route set to the destination exchange is determined (see Section 1.3). The call then moves to *selecting route state*, in which SSP attempts to find an available trunk in one of the trunk groups belonging to the specified route set.

Triggers can be encountered in the originating null, the collecting information, and analyzing information states. If this happens, call processing is suspended, and the call moves into the *wait* state.

The trigger types, which are shown along the transitions to the wait state, are discussed later. When the call enters the wait state, SSP orders its AIN-ASE to initiate a transaction with SCP, to obtain call-handling instructions. The message from SCP that ends the transaction includes one of the following instructions.

Analyze Route (ANR). This message instructs SSP to resume the call processing at the analyzing information state, or the selecting route state, depending on a parameter in the message (to be discussed later).

Disconnect (DISC). This instruction is sent when SCP has determined that the call should not be set up, and that the calling party should receive a standard tone (say, reorder-tone) or announcement. SSP then resumes the call processing at the *disconnect* state. SSP connects the calling line or incoming trunk to a tone or announcement circuit, and the caller hears a standard tone (e.g., reorder-tone) or an announcement. When SSP receives a disconnect signal from the line or the trunk, it clears the connection.

Send to Resource, Final Announcement (STR-F). This instruction is sent when SCP has determined that the call should not be set up, and that the calling line should be connected to an INSC (resource), which should play the AIN announcement specified in the instruction. SSP then resumes the processing at the *final announcement* state. It connects the calling line or incoming trunk to an INSC, plays the specified announcement, and awaits the disconnect signal from the line or trunk.

Send to Resource, Caller Interaction (STR-I). This instruction indicates that SCP needs to interact with the calling party. It requests SSP to resume processing at the *caller interaction* state. SSP connects the calling line or incoming trunk to an INSC and plays the specified announcement. The announcement contains a question and prompts the calling party to respond with one or more digits. The INSC collects the digits, and SSP sends them to SCP in a Resource Clear message. The call reenters the wait state and stays there until a new message with instructions from SCP has been received.

Terminating BCM (T-BCM). This consists of the states of calls with called party numbers that identify SSP as the terminating exchange—see Fig. 17.2-2. The calls may have originated by a customer served by SSP or may have reached SSP on an incoming trunk.

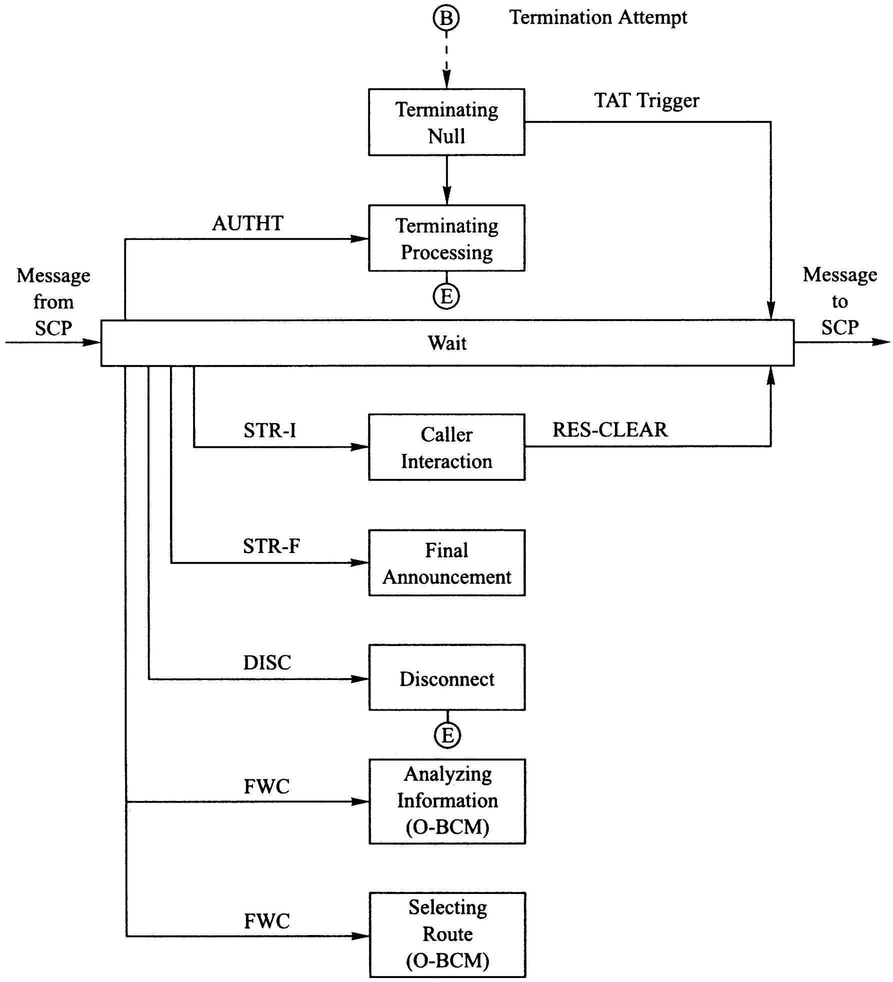


Figure 17.2-2. Terminating BCM.

The call model is entered at point B (*terminating null* state). If no trigger is encountered, the call moves to the *terminating processing* state. If a trigger has been encountered, the call moves to the wait state, and SSP initiates a transaction with SCP. If SCP determines that the call destination is at SSP, it returns an Authorize Termination (AUTHT) message, and SSP resumes processing, reentering the terminating processing state.

If SCP determines that the call destination is at another exchange, it returns a Forward Call (FWC) message. SSP then resumes call processing at the analyzing information state, or the selecting route state, of the originating BCM.

The SSP actions on receipt of a Disconnect (DISC), a Send to Resource, Final (STR-F), or a Send to Resource, Interactive (STR-I) message are the same as in the originating BCM. However, depending on where the call originated, the standard tone or announcement source or INSC is now connected to the calling line or to the incoming trunk.

17.2.2 Trigger Types and Call States

We begin our examination of the various triggers by listing the call states in which they can be encountered—see Figs. 17.2-1 and 17.2-2.

Off-Hook Immediate (OHI) Trigger. This trigger can be encountered when a call moves out of the originating null state (i.e., when a customer served by SSP goes off-hook).

Off-Hook Delay (OHD) Trigger. This trigger can be encountered in the collecting information state, when the digits from a calling customer or from a private facility trunk have been received.

Feature Code Dialing (FCD) Trigger. This trigger can be encountered in the analyzing information state, if the caller has dialed a *feature code* of the public office dialing plan.

Customized Dialing Plan (CDP) Trigger. This trigger can be encountered in the analyzing information state, if the caller using a customized dialing plan has dialed an *access code* belonging to the plan.

3/6/10 Digit (3/6/10D) Triggers. This trigger can be encountered in the analyzing information state, if the caller has dialed a called number of the NANP.

Termination Attempt (TAT) Trigger. This trigger can be encountered when a call leaves the terminating null state, and the digits received from the calling customer or incoming trunk represent the NANP number of a customer served by SSP.

17.2.3 Subscribed, Group, and Office-Based Triggers

The OHI, OHD, FCD, and TAT triggers are *subscribed* triggers. They are assigned to the lines, private facility groups, and/or directory numbers on customer request. The customers are charged a monthly or a per-use fee.

The CDP triggers are *group* triggers and are assigned, on request, to the lines of a business group. They also require a fee.

3/6/10 Digit triggers are *office-based*. They do not involve charges to the customers whose calls encounter them.

17.2.4 Encountering a Trigger

A SSP encounters a trigger (and initiates a transaction with SCP) when the following conditions are met:

1. The trigger is either:
   - Office based, or
   - Group based, and the call is originated by a line or a private facility trunk associated with a business group, or
   - Subscribed, and the call is originated by a line that has subscribed to the (OHI, OHD, FCD) trigger or terminates at a called number that has subscribed to the TAT trigger.
2. The trigger is active. The OHI, OHD, FCD, CDP, and TAT triggers can be activated and deactivated on command from SCP.
3. The criteria for the trigger type have been met.

We now examine the assignments and criteria for the various triggers.

17.2.5 Off-Hook Immediate Trigger

Off-hook immediate (OHI) triggers can be assigned to analog lines attached to SSP. When the line goes off-hook, the trigger is encountered, and SSP immediately initiates a transaction.

17.2.6 Off-Hook Delay Trigger

Off-hook delay (OHD) triggers can be assigned to analog or digital (ISDN) subscriber lines, or private facility trunks, attached to SSP. The trigger is detected when an analog subscriber line goes off-hook, a setup message is received from an ISDN line, or a seizure signal has been received on a private facility trunk.

The primary criterion for OHD is that the customer is making a call. This can be determined only after the customer's dialed digits have been received.

If the customer has dialed a valid called number, say, AC(3)-*NXX-XXXX* or *NXX-XXXX*, the OHD trigger is usually encountered. However, a telecom can designate certain called numbers as "escape codes." A call to an escape code—for example, 911 (emergency)—escapes the OHD trigger, and no transaction is initiated.

Next, suppose that the customer has dialed an access code to an exchange-based feature, followed by a called party number. This may—or may not—mean that a call is being originated. For example, $*$67-*NXX-XXXX* indicates that the customer is originating a call and does not want his number to be presented to the called party (Section 3.7.1). In this case, the OHD trigger is encountered, unless the called number is an escape code. However, the digit string 74#-3-*NXX-XXXX* is a request by a customer—who has "speed calling" service—to change the third entry on his speed-calling list to *NXX-XXXX*. Since no call is originated, the OHD trigger is not encountered.

17.2.7 Feature Code Dialing Trigger

The feature code dialing (FCD) trigger can be assigned to analog and digital subscriber lines and to private trunk groups. This enables them to invoke an AIN service, by dialing a FCD feature code.

In addition, one of the following criteria, which depend on the particular FCD code, has to be met.

- No digits follow the FCD code.
- The digits following the FCD code represent a valid NANP number.
- The digits following the FCD code represent a number of another type, say, an authorization code or a personal identification number (PIN).

The FCD codes are considered as belonging to the "public" dialing plan. On calls originated by a line or private facility trunk using a customized dialing plan, SSP considers the received digits as CDP numbers and does not recognize FCD codes. However, a calling CDP customer can "escape" to the public numbering plan, by dialing an access code in the custom dialing plan—say, $*9$.

The caller then receives a second dial-tone, and SSP interprets the subsequent received digits as belonging to the public dialing plan. This enables the CDP customer to invoke the AIN services in the manner described above, provided that a FCD trigger has been assigned the line or trunk.

17.2.8 Custom Dialing Plan Trigger

Custom dialing plan (CDP) triggers can be assigned to lines and private facility trunks of business customer groups using a customized dialing plan. In these plans, the caller dials a called number (in the particular CDP format), or an *access code* defined for the CDP, possibly followed by additional digits. Some access codes invoke exchange-based services. Other codes invoke AIN services that have been defined for the business group and require the assignment of a CDP trigger. One of the following criteria, which depend on the particular access code, has to be met:

- No digits follow the access code.
- The digits following the access code represent a valid number in the CDP of the business group.
- The digits following the access code represent a number of another type, say, an authorization code or a personal identification number (PIN).

17.2.9 3/6/10 Digit Trigger

The telecom can assign these triggers to the first three, the first six, or all digits of NANP numbers:

AC(3)	Area code (three digits)
AC(3)-*NXX*	Area and exchange code (six digits)
AC(3)-*NXX-XXXX*	Complete national number
SAC	Service access code (700, 800, 900, 976, etc.)
SAC-*NXX*	First six digits of a 700, 800, 900, etc. number
SAC-*NXX-XXXX*	Complete (ten-digit) 700, 800, 900, etc. number

The criterion for this trigger is met when a call is made to any NANP number to which the trigger has been assigned. Calls made by any originator (line, private facility trunk) using the NANP can encounter the trigger.

17.2.10 Terminating Attempt Trigger (TAT)

Terminating attempt triggers (TATs) can be assigned to directory numbers covered by the SSP exchange. Any TAT detected during processing in the terminating null state is considered to be encountered.

17.3 AIN MESSAGES AND TRANSACTIONS

This section explores the most important messages and transactions of AIN 0.1 [16,17,20].

17.3.1 Messages

AIN messages are grouped into a number of message families—see Table 17.3-1.

Request Instructions Family. The messages are sent by SSP when a trigger has been encountered and request SCP to provide instructions for handling the call.

The names of the messages in this family represent the SSP event during which the trigger has been encountered—see Figs. 17.2-1 and 17.2-2.

Origination_Attempt. This message requests instructions for a call that has encountered a trigger (OHI) in the originating null state of O-BCM.

Info_Collected. This message requests instructions for a call that has encountered a trigger (OHD) in the collecting information state of O-BCM.

Info_Analyzed. This message requests instructions for a call that has encountered a trigger (FCD, CDP, or 3/6/10D) in the analyzing information state of O-BCM.

TABLE 17.3-1 AIN 0.1 Messages

	Sender		Package			Component	
Message	SSP	SCP	QRY	CON	RES	INV	RR
Request Instructions Family							
Origination_Attempt	X		X			X	
Info_Collected	X		X			X	
Info_Analyzed	X		X			X	
Termination_Attempt	X		X			X	
Connection Control Family							
Analyze_Route		X			X	X	
Authorize_Termination		X			X	X	
Disconnect		X			X	X	
Send_To_Resource_Final		X			X	X	
Connectivity Control Family							
Forward_Call		X			X	X	
Caller Interaction Family							
Send_To_Resource_Interaction		X		X		X	
Resource_Clear	X			X		X	
Status Notification Family							
Monitor_For_Change		X	X			X	
Monitor_For_Success	X			X		X	
Status_Reported	X				X	X	
Information Revision Family							
Update_Request		X	X	X		X	
Update_Data	X				X		X

Note: QRY, query package; CON, conversation package; RES, response package; INV, invoke component; RR, return-result component.

Source: TR-NWT-001285. Reprinted with permission of Bellcore. Copyright © 1992.

Termination_Attempt. This message requests instructions for a call that has encountered a trigger (TAT) in the null state of T-BCM.

Connection Control Family. These messages are sent by SCP and contain call-control instructions.

Analyze_Route. This message instructs SSP to resume call processing, at the analyzing information or the selecting route state of O-BCM (the choice depends on parameters in the message).

Authorize_Termination. This message instructs SSP to resume processing at the terminating call processing state of T-BCM.

Disconnect. This message instructs SSP to resume processing the call at the disconnect state of O-BCM or T-BCM. SSP then connects the calling line or incoming trunk to a (non-AIN) tone or announcement source, to inform the calling party that the connection cannot be set up.

Send_To_Resource_Final. This message instructs SSP to resume the processing at the final announcement state of O-BCM or T-BCM. SSP then connects the calling line or incoming trunk to an INSC and plays the announcement specified by SCP.

Connectivity Control Family. This family consists of just one message, which is sent by SCP.

Forward_Call. This message instructs SSP to resume the processing at the analyzing information state, or the selecting route state, of O-BCM.

Caller Interaction Family. This family consists of two messages.

Send_To_Resource_Interaction. This message is sent by SCP and requests SSP to resume the processing at the caller interaction state in O-BCM or T-BCM. SSP then connects the calling line or incoming trunk to an INSC (the resource), plays an announcement specified by SCP, and collects the digits sent by the calling party.

Resource_Clear. This message is sent by SSP, after it has collected the digits received by INSC and has released the resource. The message includes the collected digits.

Status Notification Family. This family consists of three messages.

Monitor_For_Change. This message is sent by SCP and requests SSP to determine the status (idle or busy) of a specified *facility* (a line, a group of lines, or a public or private trunk group).

Monitor_For_Success and Status_Reported. These messages arc sent by SSP, in response to a monitor for change message (Section 17.3.4).

Information Revision Family. These messages enable SCP to activate or deactivate an OHI, OHD, FCD, CDP, or TAT trigger at a SSP.

Update_Request. This message is sent by SCP, and requests SSP to change the status of a specified trigger.

Update_Data. This message is sent by SSP, in response to an update request.

In addition to listing the messages and the message senders, Table 17.3-l shows that, with one exception, all messages are invokes. It also lists the TCAP packages (query, conversation, response) that can contain the various messages.

We next explore a number of AIN transactions. We distinguish call-related transactions (initiated by SSP when a trigger has been encountered), and transactions that are not call-related (and initiated by SCP).

17.3.2 Basic Call-Related Transactions

A basic call-related transaction involves a TCAP *query* package sent by SSP, containing a message of the request instructions family, and a *response* package, sent by SCP and containing a message of the connection control, or connectivity control, family.

Figure 17.3-1(*a*) shows basic transactions that are initiated because a trigger has been encountered in the originating BCM. The query and response packages are denoted by Q and R. Their messages are shown inside square brackets and, in query packages, the acronyms of the triggers are shown in parentheses. For example, a query package with an Info_Analyzed message can be the result of encountering an FCD, CDP, or 3/6/10D trigger.

Figure 17.3-2 shows basic transactions initiated as a result of encountering a TAT trigger (in the terminating BCM).

17.3.3 Interactive Call-Related Transactions

Any of the transactions described above can include one or more pairs of conversation (C) packages. In the example of Fig. 17.3-2, SCP has received a query package

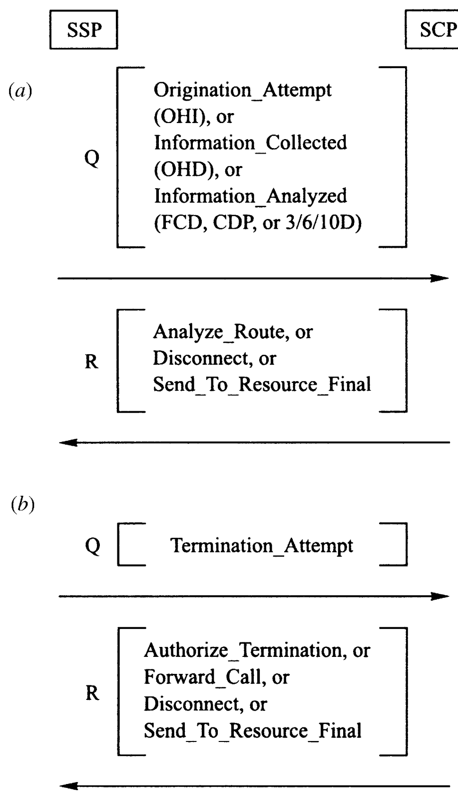


Figure 17.3-1. Basic call-related transactions. Q, Query package; R, response package.

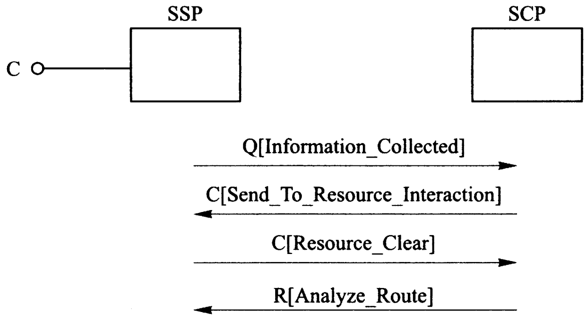


Figure 17.3-2. Call-related transaction with caller interaction. Q, Query package; C, conversation package; R, response package.

with an Info_Collected message and returns a conversation package with a Send_To_Resource_Interaction (STR-I) message.

After connecting the calling line or incoming trunk to an INSC, playing the specified announcement, and collecting the caller's DTMF digits, SSP returns a conversation package with a Resource_Clear (RES-CLR) message, which includes the collected digits.

The SCP uses this information to formulate the call-handling instructions and ends the transaction with a response package, which in this example contains an Analyze_Route message.

17.3.4 Status Monitoring Transaction

The status monitoring transaction is not call-related and enables a SCP to determine the status (busy or idle) of a "facility" (line or trunk) or a "facility group" (line group or trunk group) attached to a SSP.

In the example of Fig. 17.3-3, SCP needs to know the status of a line and initiates the transaction with the SSP to which the line is attached. The query package holds a Monitor_For_Change message, which specifies the line to be monitored, and a line state. In case (a), SSP finds that the line is in the state specified by SCP. It returns a response package with a Status_Reported message. This ends the transaction.

In case (b), SSP finds that the line is not in the specified state. It then returns a conversation package with a Monitor_Success message and keeps monitoring the line. When the line changes to the specified state, SSP ends the transaction, sending a response package with a Status_Reported message.

17.3.5 Trigger Activation and Deactivation

This transaction enables a SCP to activate or deactivate a trigger in a SSP. When SCP has determined that the status of a trigger has to be changed, it opens a transaction with the SSP where the trigger is located—see Fig. 17.3-4. The query package contains

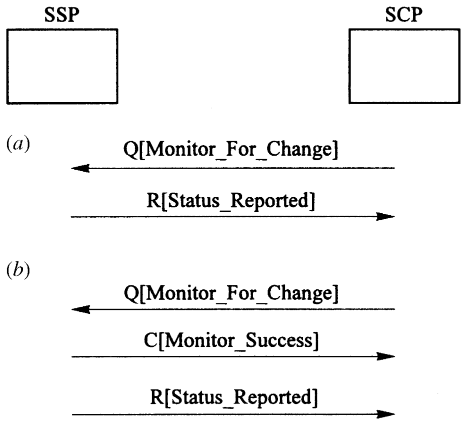


Figure 17.3-3. Status monitoring transactions.

An Update_Request message, in which the trigger and its activity status are specified. After executing the operation, SSP returns a response package. This package is somewhat of an exception: the Update_Data message is not an invoke, but a return-result.

17.3.6 SCCP Addresses

Call-related transactions are initiated by the intelligent network (IN) ASE at a SSP. For a query package (which is the first package of the transaction), ASE provides a global title (GT) called address (Section 15.3.3). The meaning of the GTA (address) part of GT (ten BCD digits) depends on the type of encountered trigger. For queries resulting from OHI, OHD, FCD, and CDP triggers, GTA is the calling customer's charge number. For the TAT and 3/6/10D triggers, GTA is the called party number.

The translation type (TT) of GT specifies that the number in GTA is to be translated into the PC-SSN address of the IN-ASE in a SCP that stores information for the number.

Transactions that are not call-related are initiated by a SCP. The GTA of the global title is a ten-digit number that represents the line, multiline group, or private facility group to be monitored, or the line or private facility trunk to

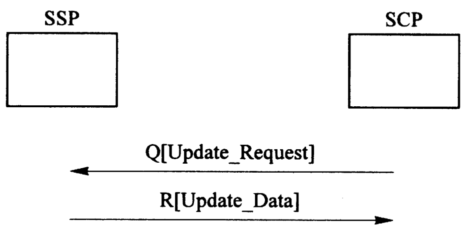


Figure 17.3-4. Information revision transaction.

which the trigger to be activated/deactivated is assigned. The TT value indicates that GTA is to be translated into the PC-SSN address of the IN-ASE in a SSP.

In the subsequent messages of a transaction, a GT translation is not necessary, because the ASEs then know each other's PC-SSN address.

17.4 AIN 0.1 PARAMETERS

This section outlines the most important parameters in AIN 0.1 messages [17,20]. At this point, it is suggested that these descriptions are merely perused. Each parameter has a reference number (e.g., Par.l). This enables the reader to quickly locate a parameter description when reading the material in the later sections of this chapter.

17.4.1 Introduction

Before describing the parameters, we need to make a few introductory remarks.

Parameter Names. The names of most parameters consist of several words. This section follows the AIN convention for denoting parameter names. Multiword names are written without spacings, and with the first letter of each word in uppercase, for example, CallingPartyID.

Primitives and Constructors. Most parameters are "primitive" data elements (Section 16.3.1). However, Par.l (UserID) and Par.29 (STRParameterBlock) are "constructor" DEs.

Contents Fields. The contents fields of primitive parameters mostly consist of octets holding an integer and octets holding two BCD digits.

The format of the contents field shown in Fig. 17.4-1, which is known as the *AIN-digits* format, has been carried over from ISUP signaling (see Fig. 11.2-7). It is used for several AIN parameters that represent "numbers" coded as BCD digits—for example, the CallingPartyID.

Bit O/E indicates whether the number of digits is even (0) or odd (1). The digits are in octets $2-n$ (if the number of digits is odd, bits 8 to 5 of octet n contain a filler code).

The *numbering plan* (NP) field indicates whether the number belongs to a *public* (001) or *private* (101) numbering plan. In this context, "public" means the PSTN/ISDN numbering plan of ITU-T Recommendation E.164 [21] or, more specifically, the North American numbering plan NANP (Section 1.2.1). Private numbers are numbers belonging to a customized dialing plan (CDP).

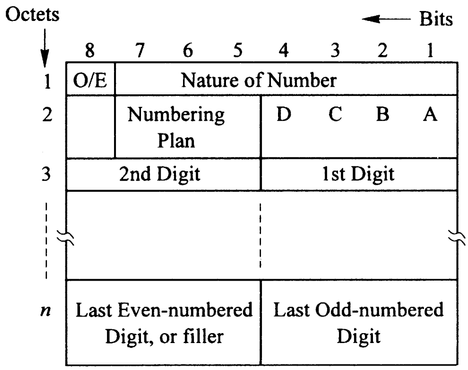


Figure 17.4-1. AINDigits format. (From Rec. Q.763. Courtesy of ITU-T.)

17.4.2 Parameter Descriptions

Par.1 UserID. This parameter identifies the "user" of the encountered trigger. It is a "constructor"—see Fig. 17.4-2. Its value (contents) field V1 holds another parameter, for example:

Par.2	DirectoryNumber (DN)
Par.3	TrunkGroupID
Par.4	PrivateFacilityGID

In queries caused by OHI, OHD, or FCD triggers, UserID identifies a calling entity that is using the public numbering plan to which the encountered trigger is assigned.

In queries resulting from CDP triggers, UserID identifies the calling entity (using a private numbering plan) to which the encountered trigger is assigned.

In queries caused by a TAT trigger, UserID identifies the called customer to whom the trigger is assigned.

In queries with 3/6/10D triggers, UserID identifies the calling entity that has dialed a NANP number that caused an encounter with a 3-, 6-, or 10-digit trigger.

Par.2 DirectoryNumber (DN). The five-octet contents field of this parameter holds ten BCD digits that represent the NANP number of a customer.

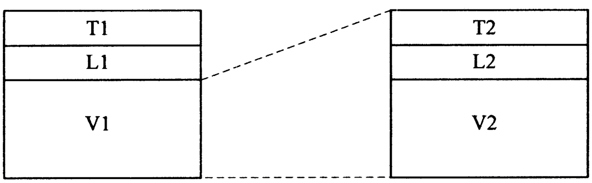


Figure 17.4-2. Format of UserID.

Par.3 TrunkGroupID. This is an integer that specifies a public trunk group.

Par.4 PrivateFacilityGID. This is an integer that specifies a private facility trunk group.

Par.5 BearerCapability. This parameter indicates the bearer capability requested by the calling party (Section 10.3.5). The contents field is coded as follows:

0000 0000	Speech
0000 0001	3.1-kHz Audio
0000 0011	56-kb/s Data
0000 0100	64-kb/s Data

Par.6 TriggerType. This parameter indicates the type of a trigger:

0000 0001	Feature dialing code (FCD) trigger
0000 0010	Customized dialing plan (CDP) trigger
0000 0100	Three-digit trigger (area code)
0000 0101	Six-digit trigger (area and exchange code)
0000 1000	Ten-digit trigger (national number)
0000 1111	Termination attempt (TAT) trigger
0001 0000	Off-hook immediate (OHI) trigger
0001 0001	Off-hook delay (OHD) trigger

Par.7 CallingPartyID. This parameter represents the number of the calling party. Its contents field has the AINDigits format.

Numbering Plan (NP). This can be public or private.

Nature of Number (NN)

000 0001	Subscriber number
000 0011	National number
000 0100	International number

Bits BA. These indicate whether the calling customer allows the presentation of his number to the called party:

00	Presentation allowed
01	Presentation restricted

Bits DC. These indicate whether the calling number was provided by a calling ISDN user or by the exchange:

01	Provided by user, not screened by exchange
01	Provided by user, passed screening by exchange
11	Provided by exchange

Par.8 ChargeNumber. This parameter identifies the charge number of the calling customer. If the customer has one line, the charge number is usually the caller's number. If a customer has a group of lines, a common number can be designated as the charge number for all lines of the group, and the customer receives one consolidated monthly bill. The contents field has the AINDigits format.

Nature of Number (NN)

000 0001	Subscriber number (calling party)
000 0011	National number (calling party)
000 0101	Subscriber number (called party)
000 0111	National number (called party)

Numbering Plan (NP). This can be public or private.

Bits C,D,B,A. These are not used.

Par.9 LATA. This parameter represents the LATA network of the calling party (Section 1.1.2) and consists of three BCD digits. The contents field has the AINDigits format. The NN and NP fields are set to "not applicable" (all zeros).

Par.10 PrimaryCarrier. This parameter specifies the first-choice long-distance (interexchange) carrier for the call. Octet 1 of the contents field indicates whether the carrier is the regular (presubscribed) long-distance carrier of the calling customer or has been selected by the customer, by dialing a 10XXX prefix (Section 3.7.1):

0000 0000	No indication
0000 0001	Regular (subscribed) carrier
0000 0100	Dialed by calling party

Octets 2 and 3 contain the three or four BCD digit carrier identification.

Par.11 AlternateCarrier. This parameter specifies the second-choice long-distance carrier for the call. The contents field is coded as in Par.10.

Par.12 SecondAlternateCarrier. This parameter specifies the third-choice long-distance carrier for the call. The contents field is coded as in Par.10.

Par.13 CalledPartyID. This represents the number of the called party and indicates whether the calling party has requested operator assistance for the call. The contents field has the AINDigits format.

Nature of Number (NN)

Bits 7–5	
000	No assistance requested
111	Assistance requested

Bits 4–1	
0001	Subscriber number
0011	National number
0111	International number

Numbering Plan (NP). This can be public or private.

Par.14 FeatureCode (VerticalServiceCode). This is a code, dialed by a caller who is using the public numbering plan, to request an AIN service.

The contents field is in the AINDigits format. The NN and NP fields are set to "not applicable" (all zeros).

The code is usually an ∗ (asterisk), followed by up to four digits.

Par.15 AccessCode. This parameter is a code, dialed by a caller who is using a customized dialing plan, to request an AIN service. The contents field is as described for Par.14.

Par.16 CollectedAddressInfo. This parameter contains address information collected from the calling party. It has the same format and coding as Par.13 (CalledPartyID).

Par.17 CollectedDigits. This parameter holds the digits collected from the user, which may be a called party number or a number with a different meaning (e.g., authorization code, personal identification code). The contents field has the AINDigits format.

The NN and NP fields are set to "unknown" (000 0000 and 000).

Par.18 CallingPartyBGID. This parameter is included in queries, if the calling line is a member of a business group. It identifies the group and also indicates whether the line is restricted from making or receiving certain calls. The contents and coding have been carried over from U.S. ISUP signaling—see Section 11.9.2.

Par.19 PrimaryBillingIndicator. There are several “billing indicator” parameters. They contain information required by the telecom accounting centers, for the calculation of the call charge.

The contents field of these parameters have four octets, holding two BCD digits each.

Octets 1 and 2 specify the *automatic message accounting (AMA) call type*. Octets 3 and 4 contain a *service feature identifier*.

Par.19 is the billing indicator to be used if SSP routes the call to a trunk of the first-choice private trunk group.

Par.20 AlternateBillingIndicator. The meaning and contents of this parameter are as in Par.19.

Par.20 is the billing indicator to be used if SSP routes the call to a trunk of the second-choice private trunk group.

Par.21 SecondAlternateBillingIndicator. The meaning and contents of this parameter are as in Par.19.

Par.21 is the billing indicator to be used if SSP routes the call to a trunk of the third-choice private trunk group.

Par.22 OverflowBillingIndicator. The meaning and contents of this parameter are as in Par.19.

Par.22 is the billing indicator to be used if SSP has tried unsuccessfully to route the call to a private facility trunk and is now routing the call to a trunk in the public network.

Par.23 AMAAlternateBillingNumber. This parameter is included in a SCP response, if the call is to be charged to a number other than the ChargeNumber (Par.8). The contents field is coded as in Par.8.

Par.24 PrimaryTrunkGroup. This parameter specifies the first-choice private facility trunk group for the call and the digits to be outpulsed (sent out). Octets 2–5 of the contents field contain eight BCD digits that specify the group. Bit H in octet 1 indicates the parameter that holds the digits to be outpulsed:

H = 0	Par.27 (OutpulseNumber)
H = 1	Par.13 (CalledPartyID)

Par.25 AlternateTrunkGroup. This parameter specifies the second-choice private facility trunk group for the call. Its contents are coded as in Par.24.

Par.26 SecondAlternateTrunkGroup. This parameter specifies the third-choice private facility trunk group for the call. Its contents are coded as in Par.24.

Par.27 OutpulseNumber. This parameter holds the digits to be outpulsed on a private facility trunk. The contents field has the AINDigits format. The NN and NP fields are set to "not applicable" (all zeros).

Par.28 ResourceType. This parameter differentiates the "final" and "interaction" send to resource operations (STR-F and STR-I). The contents field is coded as follows:

0000 0000	Final STR operation
0000 0001	Interaction STR operation

Par.29 StrParameterBlock. This parameter specifies the announcement to be played by an INSC.

The parameter is a constructor. Its contents field holds one or more *announcement elements*—see Fig. 17.4-3. *AnnouncementID* is an integer that specifies a phrase (e.g., "please enter your authorization code"). A phrase may be followed by one or more spoken *information digits*, which are specified by BCD numbers.

For interactive operations, the element also has a *MaximumDigits* field, which is an integer that specifies the number of digits to be collected:

Integer	Number of Digits
0–31	As specified by the integer
253	"Normal number of digits"
254	"Any number of digits"

Par.30 DisconnectFlag. The presence of this parameter in a message indicates that the calling party should be disconnected after the resource (INSC) has played the announcement. The parameter has no contents field.

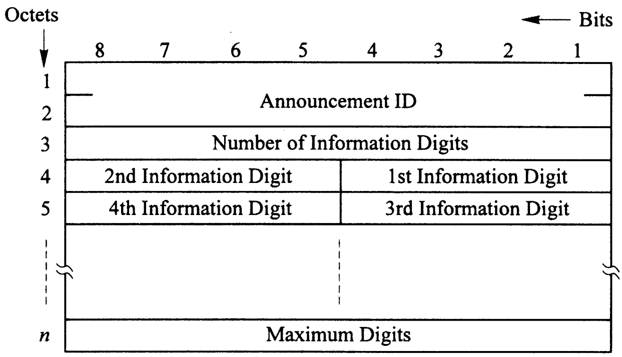


Figure 17.4-3. Announcement element. (From TR-NWT-001285. Reprinted with permission of Bellcore. Copyright © 1992.)

Par.31 ClearCause. This indicates why the SSP has cleared a resource (INSC):

0000 0000	Normal clearing
0000 0010	Timeout
0000 0110	Caller disconnect

Par.32 FacilityStatus. This parameter indicates the status of a line, a private facility, a public or private trunk group, and so on:

0000 0000	Busy
0000 0011	Idle

Par.33 FailureCause. This indicates why SSP has been unable to execute a requested operation:

0000 0001	Rate too high
0000 0010	Unavailable resources

Par.34 MonitorTime. This parameter specifies the length of time that SSP has to monitor the status of the line (or line group, etc.).

The contents field consists of three octets, each of which holds two BCD digits. Octets 1, 2, and 3 specify the number of hours, minutes, and seconds, respectively.

Par.35 StatusCause. This indicates why SSP has ended the monitoring of a line or facility group and is sending a Status_Reported message:

0000 0000	The status of the monitored line or facility group matches the specified state
0000 0001	The status does not match the specified state, and the monitoring time has elapsed

Par.36 TriggerTypeFlag. This parameter identifies a trigger type and specifies whether the trigger is to be activated. The parameter has a two-octet contents field. Octet 1 specifies the trigger status:

0000 0000	Do not activate trigger
0000 0001	Activate trigger

Octet 2 identifies the trigger type and is coded as in Par.6.

17.5 CODING OF DATA ELEMENTS

This section outlines the coding of AIN 0.1 data elements (DEs) in the components part of TCAP packages [17,20].

17.5.1 Tag Codes

All data elements other than parameter DEs are "national TCAP" DEs. In their one-octet tags (specifiers), bits H,G are set to 1,1 (Section 16.2).

The parameter DE tags are coded "context specific" (H,G = 1,0), in the context of the parameter set DE (which itself is a national TCAP DE).

The tag fields of parameter DEs consist of one or two octets. The integers in the tag code fields (Fig. 16.2-3) are listed in Table 17.5-1.

TABLE 17.5-1 AIN 0.1 Parameter Tags

Reference	Parameter	Tag (Integer)
Par.1	UserID	53
Par.2	DN (DirectoryNumber)	— [1]
Par.3	TrunkGroupID	27 [5]
Par.4	PrivateFacilityGID	28 [6]
Par.5	BearerCapability	13
Par.6	TriggerType	52
Par.7	CallingPartyID	18
Par.8	ChargeNumber	19
Par.9	LATA	35
Par.10	PrimaryCarrier	41
Par.11	AlternateCarrier	4
Par.12	SecondAlternateCarrier	47
Par.13	CalledPartyID	15
Par.14	FeatureCode	54
Par.15	AccessCode	1
Par.16	CollectedAddressInfo	22
Par.17	CollectedDigits	23
Par.18	CallingPartyBGID	17
Par.19	PrimaryBillingIndicator	40
Par.20	AlternateBillingIndicator	3
Par.21	SecondAlternateBillingIndicator	46
Par.22	OverflowBillingIndicator	38
Par.23	AMAAlternateBillingNumber	6
Par.24	PrimaryTrunkGroup	42
Par.25	AlternateTrunkGroup	5
Par.26	SecondAlternateTrunkGroup	48
Par.27	OutpulseNumber	37
Par.28	ResourceType	45
Par.29	StrParameterBlock	50
Par.30	DisconnectFlag	25
Par.31	ClearCause	21
Par.32	FacilityStatus	61
Par.33	FailureCause	34
Par.34	MonitorTime	65
Par.35	StatusCause	66
Par.36	TriggerTypeFlag	68

Note: [1], [5], and [6] are the tag values of Par.2, Par.3, and Par.4 when the parameter is held in Par.1 (UserID).

Source: TR-NWT-001285. Reprinted with permission of Bellcore. Copyright © 1992.

TABLE 17.5-2 Contents of Operation Code DEs in AIN 0.1 Messages

Message	Octet	← Bits HGFE DCBA
Request Instructions Family	1	0110 0100
Origination_Attempt	2	0001 1000
Info_Collected	2	0000 0010
Info_Analyzed	2	0000 0011
Termination_Attempt	2	0000 0101
Connection Control Family	1	0110 0101
Analyze_Route	2	0000 0001
Authorize_Termination	2	0000 0010
Disconnect	2	0000 0011
Send_To_Resource_Final	2	0000 0001
Connectivity Control Family	1	0110 1010
Forward_Call	2	0000 0001
Caller Interaction Family	1	0110 0110
Send_To_Resource_Interaction	2	0000 0001
Resource_Clear	2	0000 0010
Status Notification Family	1	0110 0111
Monitor_For_Change	2	0000 0001
Monitor_For_Success	2	0000 0011
Status_Reported	2	0000 0101
Information Revision Family	1	0110 1000
Update_Request	2	0000 0001
Update_Data	—	—

Notes: (1) Send_To_Resource_Final and Send_To_Resource_Interaction have the same operation code. The messages are differentiated by Resource Type (Par.28). (2) Update_Data is a Return-Result component.

Source: TR-NWT-001285. Reprinted with permission of Bellcore. Copyright © 1992.

17.5.2 Contents of the Operation Code DE

Operation code DEs have two-octet contents fields. Octet 1 identifies the operation family, and octet 2 represents an operation within that family—see Table 17.5-2.

17.5.3 Contents of Parameter DEs

The contents fields of parameter DEs have been outlined in Section 17.4.2. Additional details can be found in [17,20].

17.6 MESSAGES AND PARAMETERS

This section revisits the messages described in Section 17.3, this time including their most important mandatory (M) and optional (O) parameters. It is helpful to look up the parameter descriptions of Section 17.4 when reading the material that follows.

With the exception of Update_Data, the messages, invoke operations specified by an *operation code* DE, and the message parameters are the "operands" for the operation.

17.6.1 Messages in the Request Instructions Family

The messages in this family are sent by SSP and invoke SCP operations. The messages include parameters with attributes of the calling, line or trunk, which are stored in semipermanent memory of SSP, and parameters with information "dialed" by the caller. The parameters are listed in Table 17.6-1.

For calls originated by analog or ISDN lines attached to SSP, UserID (Par.l), Charge Number (Par.8), LATA (Par.9), and CallingPartyBGID (Par.18) are stored attributes.

Some parameters are stored attributes of the calling analog subscriber line or the incoming non-ISUP trunk, but are parameters in SETUP messages received from ISDN users (Section 10.3.2) or in IAM (Initial Address) messages for trunks with ISUP signaling (Section 11.2.4). For example, BearerCapability (Par.5) for analog lines always indicates "3.1-kHz audio," but the BearerCapability received in SETUP and IAM messages can indicate various capabilities. The same holds for CallingPartyID (Par.7).

PrimaryCarrier (Par.10) is an attribute of analog and ISDN lines that can be overwritten by the caller, by dialing a 10*XXX* code.

CalledPartyID (Par.13), FeatureCode (Par.14), AccessCode (Par.15), Collected-AddressInfo (Par.16), and CollectedDigits (Par.17) are parameters that always hold information received from the calling party. Table 17.6-2 shows the combinations of these parameters in messages caused by the various trigger types. For example, the

TABLE 17.6-1 Parameters in Messages of the Request Instructions Family

		Message			
Reference	Parameter	(1)	(2)	(3)	(4)
Par.1	UserID	M	M	M	M
Par.5	BearerCapability	M	M	M	M
Par.6	TriggerType	O	O	O	O
Par.7	CallingPartyID	O	O	O	O
Par.8	ChargeNumber	O	O	O	O
Par.9	LATA	O	O	O	O
Par 10	PrimaryCarrier	O	O	O	
Par.13	CalledPartyID			O	O
Par.14	FeatureCode		O	O	
Par.15	AccessCode		O	O	
Par.16	CollectedAddressInfo		O	O	
Par.17	CollectedDigits		O	O	
Par.18	CallingPartyBGID			O	

Messages: (1) Origination_Attempt. (2) Info_Collected. (3) Info_Analyzed. (4) Termination_Attempt.

Source: TR-NWT-00 1285. Reprinted with permission of Bellcore. Copyright © 1992.

TABLE 17.6-2 Combinations of Parameters Holding Dialed Digits

		Trigger Type			
Reference	Parameter	(1)	(2)	(3)	(4)
Par.13	CalledPartyID				H
Par.14	FeatureCode	BCD	BCD		
Par.15	AccessCode			EFG	
Par.16	CollectedAddressInfo	A D	D	G	
Par.17	CollectedDigits	C	C	F	

Trigger types: (1) Off-hook delay (OHD). (2) Feature code dialing (FCD). (3) Customized dialing plan (CDP). (4) 3/6/10 Digits and termination attempt (TAT).

Info_Collected message, sent when an OHD trigger has been encountered, may include CollectedAddressInfo only (A), FeatureCode only (B), FeatureCode and Coltected-Digits (C), or FeatureCode and Collected AddressInfo (D).

17.6.2 Messages in the Connection and Connectivity Control Families

These messages are sent by SCP and contain call-handling instructions for SSP. The message parameters are listed in Table 17.6-3.

TABLE 17.6-3 Parameters in Messages of the Connection Control and Connectivity Control Families

		Message				
Reference	Parameter	(1)	(2)	(3)	(4)	(5)
Par.7	CallingPartyID	O	O	O		
Par.8	ChargeNumber	O	O			
Par.10	PrimaryCarrier	O	O			
Par.11	AlternateCarrier	O	O			
Par.12	SecondAlternateCarrier	O	O			
Par.13	CalledPartyID	O	O			
Par.19	PrimaryBillingIndicator	O	O	O	O	O
Par.20	AlernateBillingIndicator	O	O			
Par.21	SecondAlternateBillingIndicator	O	O			
Par.22	OverflowBillingIndicator	O	O			
Par.23	AMAAlternateBillingNumber	O	O	O	O	O
Par.24	PrimaryTrunkGroup	O	O			
Par.25	AlternateTrunkGroup	O	O			
Par.26	SecondAlternateTrunkGroup	O	O			
Par.27	OutpulseNumber	O	O			
Par.28	ResourceType					O
Par.29	StrParameterBlock					O
Par.30	DisconnectFlag					O

Messages: (1) Analyze_Route. (2) Forward_Call. (3) Disconnect. (4) Authorize_Termination. (5) Send_To_Resource, Final.

Source: TR-NWT-001285. Reprinted with permission of Bellcore. Copyright © 1992.

Analyze_Route and Forward_Call. If SCP determines that a call setup can proceed, it ends the transaction with one of these messages, depending on the message received from SSP (see Fig. 17.3-1). The messages contain routing and charging information for the call.

If SCP does not provide other routing information, SSP resumes the call processing at the analyzing-information state (Fig. 17.2-l) and analyzes the called number. If the number represents a line served by SSP, the call moves to the termination_ attempt state of T-BCM. Otherwise, SCP determines the route set for the call destination. On long-distance calls, it takes into account the IC carrier selected by the calling party.

SCP can include AlternateCarrier (Par.11) and possibly SecondAlternate Carrier (Par.12). In this case, SSP also resumes processing at the analyzing-information state. If the call cannot be routed to the caller's selected IC carrier, SSP routes the call to the alternate, or second alternate, IC carrier.

If the Analyze_Route message includes PrimaryTrunkGroup (Par.24), SSP resumes, its call processing at the selecting-route state, bypassing the determination of the (public) route set for the call. This is because Par.24 specifies a private facility trunk group, owned or leased by a business customer (Section 1.1.2), as the first-choice trunk group for the call. If Par.24 is present, the message may also include Par.25 and Par.26 (second-choice and third-choice private facility trunk group), and Par.10, Par.11, and Par.12 (first-, second-, and third-choice IC carrier). SSP tests these trunk groups for available trunks, in the order listed above.

We now turn to the charging information in the Analyze_Route and Forward_Call messages. If the call is to be charged to a party other than the calling party, the charge number of this party is identified by AMAAlternateBillingNumber (Par.23). For example, in 800 calls, Par.23 holds the charge number of the business customer who "owns" the 800 number.

Par.19, Par.20, Par.21, and Par.22 (primary, alternate, second alternate, and overflow billing indicator) contain billing information to be included in the billing record for the call, if the call is routed over, respectively, the first-, second-, or third-choice private facility trunk group, or a public (IC carrier) trunk group.

Authorize_Termination Message. This message is sent when SCP has determined that the call destination is a line attached to SSP, and that the setup can proceed. Routing instructions are not needed, but the message can include charging information (Par.19 and/or Par.23).

Disconnect and Send_To_Resource (Final) Messages. These messages are sent when SCP has determined that the call should not be setup.

Both messages can include charging information (Par.19, Par.23). In Send_To_ Resource messages, ResourceType (Par.28) is set to "play announcement," and the announcement is specified by StrParameterBlock (Par.29). DisconnectFlag (Par.30) instructs SSP to disconnect the caller, after the announcement has been played.

17.6.3 Messages in the Caller Interaction Family

This family includes two messages. The message parameters are listed in Table 17.6-4.

Send_To_Resource (Interaction) Message. This message is sent by SCP, when it needs to interact with the calling party.

ResourceType (Par.28) is set to "interaction STR operation." StrParameterBlock (Par.29) contains one or more announcement elements, which specify a spoken phrase to be played, and the number of DTMF digits to be collected from the caller.

Resource_Clear Message. If the digit collection is successful, ClearCause (Par.31) indicates "normal clearing." If SCP requested address information, the collected digits are in CollectedAddressInfo (Par.16); otherwise, they are in CollectedDigits (Par.17).

If the digit collection fails, ClearCause indicates "timeout" (the expected number of digits has not been received within, say, 5 s) or "caller disconnect."

If no intelligent network service circuit (INSC) is available, SSP sends a Resource_Clear message, in which FailureCause (Par.33) indicates "unavailable resources."

17.6.4 Multiple Trigger Encounters

More than one trigger can be encountered during the processing of a call. For example, suppose that a call attempt is made from a line with an OHD trigger, and to a 900 number. The call first encounters the OHD trigger (in the collecting-information state) and—assuming that 900 calls from the line are allowed—next encounters a three-digit trigger, in the analyzing-information state (Fig. 17.2-1).

Now suppose that a call has encountered a trigger in the analyzing-information state. In most cases, SCP instructs SSP to resume call processing at the same state. In certain situations, this can cause another encounter with the same trigger.

TABLE 17.6-4 Parameters in Messages of the Caller Interaction Family

		Message	
Reference	Parameter	(1)	(2)
Par.16	CollectedAddressInfo		O
Par.17	CollectedDigits		O
Par.28	ResourceType	M	O
Par.29	StrParameterBlock	M	
Par.31	ClearCause		M
Par.33	FailureCause		O

Messages: (1) Send_To_Resource_Interaction. (2) Resource_Clear.

Source: TR-NWT-001285. Reprinted with permission of Bellcore. Copyright © 1992.

SSP programs include defenses against entering a loop that keeps on encountering a trigger. For example, if SSP has encountered a predetermined maximum number of triggers in a call, subsequent triggers are ignored.

17.6.5 Messages in the Status Notification Family

Status monitoring is a transaction in which a SCP requests a SSP to monitor the status (busy, idle) of a "facility" (line, trunk)—see Fig. 17.3-3. The parameters in the messages for this transaction are listed in Table 17.6-5.

Monitor_For_Change Message. This message, sent by SCP, initiates the transaction. CalledPartyID (Par.13) specifies the line to be monitored. The state to be monitored is indicated by FacilityStatus (Par.32). MonitorTime (Par.34) sets a limit on the monitoring time.

Status_Reported Message. This message ends a monitoring transaction. If the line is initially in the specified state, SSP immediately returns this message, and StatusCause (Par.35) indicates that the line status matches the specified state.

The SSP also immediately returns a Status_Reported message if it is unable to perform the monitoring operation. The message then includes FailureCause (Par.33). If SSP has determined that the CallingPartyID is invalid, FailureCause indicates "resource unavailable." If SSP is unable to monitor because of processor overload, FailureCause indicates "rate too high."

Monitor_For_Success Message. This message is an immediate reply to SCP, in the case that the line is initially not in the specified state.

The transaction continues, and SSP keeps monitoring the line, until either the line has changed to the specified state or the specified monitoring time has elapsed, whichever happens first. SSP then sends a (delayed) Status_Reported message in which StatusCause indicates either a "status match" or a "timeout."

TABLE 17.6-5 Parameters in Operations of the Status Notification and Information Revision Families

		Message				
Reference	Parameter	(1)	(2)	(3)	(4)	(5)
Par.1	UserID				M	
Par.13	CalledPartyID	O				
Par.32	FacilityStatus	M	M	O		
Par.33	FailureCause			O		O
Par.34	MonitorTime	M				
Par.35	StatusCause			O		
Par.36	TriggerTypingFlag				O	

Messages: (1) Monitor_For_Change. (2) Monitor_Success. (3) Status_Reported. (4) Update_Request. (5) Update_Data.

Source: TR-NWT-001285. Reprinted with permission of Bellcore. Copyright © 1992.

Snapshot Monitoring. If SCP needs a snapshot of the line status, it sets MonitorTime to "zero." SSP then immediately sends a Status_Reported, in which StatusCause indicates either a "status match" or a "timeout."

17.6.6 Messages in the Information Revision Family

These messages enable a SCP to activate or deactivate a trigger in a SSP. The message parameters are listed in Table 17.6-5.

Update_Request Message. This message is sent by SCP. TriggerTypeFlag (Par.36) specifies the trigger type and the state of the trigger. In a SSP, a trigger is uniquely identified by the combination of its type and its UserID (Par.l).

Update_Data Message. This message is the SSP response to an Update_Request. If SSP has honored the request, the message does not contain a parameter. If SSP has been unable to set the trigger to the specified state, the message includes Failure-Cause (Par.33).

17.7 AIN SERVICES

Compared with conventional "switch-based" services, the SCP-based AIN services can be designed with much more flexibility.

In the first place, the service creation environment (SCE) systems usually allow the telecom to define services that take into account the calling line—which is identified by UserID (Par.1), CallingPartyID (Par.7), or ChargeNumber (Par.8) in all messages of the request instructions family—and the time-of-day and day-of-week that the call is made.

Caller interaction also increases the flexibility of services. In the first place, services can be designed such that the calling party can select a particular option from a service menu. Moreover, a service can be made available to "authorized" callers only, by requiring the caller to enter a telecom-provided authorization code, which is checked by SCP.

This section outlines a number of services that can be implemented with the triggers, messages, and parameters that have been discussed in this chapter [20]. The examples have been chosen as illustrations for the uses of the various trigger types and messages, and regardless of their availability in U.S. networks. They are grouped into a number of categories (call management, number translation, etc.).

17.7.1 AIN Outgoing Call Management

Outgoing call management services enable residential and business customers to restrict the calls made on their lines. Some exchange-based features of this nature are already in existence. As an example, a customer can specify that calls to 900/976 numbers are to be blocked. The service is inflexible in that all 900/976 calls are either blocked or allowed.

Call Screening. Call screening is a service that enables individual customers to specify a set of numbers that should be blocked when dialed from their lines. The set may contain groups of numbers, say, all 976 numbers or all 011 (international) numbers, and individual numbers.

To make the service available, the telecom assigns an off-hook delay (OHD) trigger to the line of the customer and enters a list with dialing restrictions for this line into SCP semipermanent memory.

When the line originates a call, the OHD trigger is encountered after the dialed digits have been collected (except when the customer has dialed an escape code (Section 17.2.6).

The Info_Collected message sent by SCP identifies the calling party and includes TriggerType (Par.6) set to "OHD," and CollectedAddressInfo (Par.16), containing the called number.

Based on this information, SCP either allows or blocks the call. If the call is allowed, SCP returns an Analyze_Route message. If not, the SCP response is a disconnect or a Send_To_Resource_Final message.

Scheduled Call Screening. In this service, screening is in effect on a predetermined schedule, specified by the customer. For example, calls are screened before 8 am and after 7 pm on weekdays, and all day on weekends.

Call Screening with Override. In this service, the call screening of a line can be overridden for calls made by an "authorized" person, who identifies himself (herself) by dialing a feature code defined for the service, say, $*921$, followed by an authorization code (AC) that has been provided by the telecom.

When SSP receives a call without a feature code, the OHD trigger is encountered, and normal call screening is in effect. If an authorized caller dials $*921$ followed by AC, the OHD trigger is also encountered. SSP then sends an Info_Collected message, in which the calling line is identified, and TriggerType (Par.6) is set to "OHD." The message also includes FeatureCode (Par.14), set to $*921$, and CollectedDigits (Par.17), which holds the AC.

SCP checks whether the received AC is associated with the calling line.

If so, SCP returns a Send_To_Resource_Interaction message, in which StrParameterBlock (Par.29) indicates that dial-tone is to be sent to the caller, and that address digits are to be collected. When hearing the dial-tone, the caller dials the digits of the called number. When the digit collection is complete, SSP sends a Resource_Clear message, in which CollectedAddressInfo (Par.16) contains the called number, and SCP ends the transaction with an Analyze_Route message, in which CalledPartyID (Par.13) holds the called number received from SSP.

If SCP determines that an incorrect AC has been dialed, it returns a disconnect, or a Send_To_Resource_Final, message.

Call Screening with Activation/Deactivation. In this service, call screening can be activated and deactivated on command from an authorized person. The service requires the assignment of an OHD and a FCD trigger to the line.

We assume that the telecom has defined feature codes $*922$ and $*923$ to activate and deactivate the screening.

The activation and deactivation of the screening is accomplished by activating and deactivating the OHD trigger.

Suppose that the OHD trigger—and therefore the service—has been activated. To deactivate the service, the authorized person dials $*922$ followed by AC. The OHD trigger is encountered during the collecting-information state and results in an info_collected message, in which the calling line is identified. Also included are FeatureCode (Par.14), set to $*922$, and CollectedDigits (Par.17), which holds the AC.

If SCP determines that AC is associated with the calling line, it returns an Update_Request message, in which the trigger to be deactivated is specified by UserID (Par.1) and TriggerTypeFlag (Par.36). SSP deactivates the OHD trigger and terminates the transaction with an Update_Data message, which indicates that the trigger has been deactivated.

If SCP determines that the AC is not valid, it ends the transaction with a Disconnect, or a Send_To_Resource_Final, message.

Now suppose that the service is currently deactivated. To activate the service, the authorized customer dials $*922$ followed by AC. Since the OHD trigger of the line is deactivated, it cannot be encountered. However, the FCD trigger is now encountered, in the analyzing-information state, and SSP sends an Info_Analyzed message, with the same parameters as the Info_Collected message described above, except that TriggerTypeFlag now indicates that the OHD trigger has to be activated. The rest of the procedure again involves an Update_Request message from SCP, and an Update_Data message from SSP.

Hotline Telephones. A *hotline* is a telephone without a dial or keypad. Examples of hotlines are the emergency telephones along highways. When a caller lifts the handset, the exchange sets up a connection to a predetermined DN, of a police station or a nearby automobile service station. This feature is already available as an exchange-based service.

AIN hotline services can be more versatile. They require the assignment of an OHI trigger to the lines. SCP stores a list with directory numbers of service stations along the highway.

When receiving an Origination_Attempt message from one of the phones, SCP bases the selection of the directory number on the identity of the calling line. In this way, the call can be routed to the nearest service station. In addition, the directory number selection can be based on time-of-day and day-of-week.

17.7.2 AIN Called Number Translation

AIN number translation requires the assignment of an (office-based) 3/6/10 digit trigger on certain directory numbers. The triggers have to be assigned in all SSPs. When a SSP encounters the trigger, it sends an Info_Analyzed message. The TriggerType (Par.6) is set to "3, 6, or 10 digits." The calling party is identified, and the called number is in CalledPartyID (Par.13).

Translation of 800, 900, and 976 Numbers. Translation of 800 numbers into routing numbers (unpublished 10-digit NANP numbers) has been available since 1980, as an *intelligent network 1* (IN/1) service [1,2]. Today, most 800-number translations are still made by IN/1 databases. However, telecoms can move these translations to SCPs.

It is also possible to assign the translations of some 800 numbers to SCPs and leave the translations of other 800 numbers to the older databases. As an example, the split can be based on the *NXX* code in the 800-*NXX-XXXX* number. This is done by assigning six-digit triggers to the 800-*NXX* codes to be translated by SCP.

Translations of 900 and 976 numbers are usually done by SCPs and require the assignment of triggers to these numbers at the SSPs.

When SSP encounters one of these triggers, it sends an Info_Analyzed message in which TriggerType (Par.6) is set to "three, six, or ten digits," the called number is in CalledPartyID (Par.13), and the calling line is identified.

Translations of 800, 900, and other numbers can be tailored to the requirements of the individual business customers.

For example, a multilocation business that has an 800, or 900, or other number may require that incoming calls are routed to its nearest business office. Usually, the business customer specifies individual routing numbers for calls originated by lines in particular numbering plan areas (identified by NPA), or in a particular exchange areas (identified by NPA-*NXX*).

In addition, the determination of routing numbers can be made to depend on the time-of-day and day-of-week of the call.

In the Analyze_Route message from SCP, the routing number is in CalledPartyID (Par.13). Since the calls involve special billing and charging, the message also includes AMAAlternateBillingNumber (Par.23).

Personal Numbers. A personal number is a number by which a customer can be reached, regardless of his/her present location. We assume that the numbers have the format 500-*NXX-XXXX*. For this service, three-digit triggers are assigned to 500 numbers at all SSPs. For each 500 customer, SCP stores a "current address" by which the customer can be reached at this time.

On a call to a 500 number, SSP encounters the trigger and sends an Info_Analyzed message in which TriggerType (Par.6) indicates "3-digit" and CalledPartyID (Par.13) holds the 500 number. The SCP translates the number into the current address, and its Analyze_Route message includes the address in CalledPartyID (Par.13).

Customers with 500 numbers can update their current addresses when moving to a new location; for instance, after entering their respective homes, cars, offices, or after arriving at their hotels.

One way to update a current address is described below. The telecom establishes an "updating" number, say, 500-234-5678. To update the current address, the customer calls the updating number. SSP encounters the three-digit trigger and sends an info_analyzed message in which TriggerType (Par.6) indicates "3-digit" and CalledPartyID (Par.13) holds the updating number. SCP recognizes the updating number and executes a series of caller interactions. In the first and second interactions, SCP requests the caller's 500 number and authorization code (AC). It then checks whether the AC is associated with the caller's 500 number.

If this is the case, the third interaction prompts the caller to enter the new current address. The fourth interaction plays back the received address and requests the caller to dial "1" if the address is correct, and "2" if it is not. If SCP receives a "1," it updates the customer's current address and ends the transaction Send_To_Resource_Final message that indicates that the new address is now in effect. If SCP receives a "2," it asks again for the new address.

17.7.3 AIN Incoming Call Management

This service enables residential or business customers to specify the callers who are allowed to call their lines. This is important for lines that are connected to the customer's dial-up computers. The customer can specify a list of calling lines that are allowed to call the line(s), and possibly a list of authorization codes. For this service, a termination attempt trigger (TAT) is assigned to the directory number (DN) of the customer's line, in the SSP to which the line is attached, and SCP stores the lists with calling lines and authorization codes.

If SSP encounters a TAT trigger, it sends a Termination_Attempt message, with TriggerType (Par.6) set to "TAT trigger" CalledPartyID (Par.13) holding the called DN, and CallingPartyID (Par.7) identifying the calling line.

If the calling line is on the list of lines allowed to call the DN, SCP allows the termination. If the list shows that an authorization code (AC) is required, SCP initiates a caller interaction to obtain AC and allows the call if AC is associated with the calling line.

17.7.4 AIN Terminating Services

The procedure for providing terminating services to a line is very similar to incoming call management. A TAT trigger is assigned to the DN in the SSP that covers the DN. If SSP encounters the trigger, it sends a Termination_Attempt message with Par.6 set to "TAT" and the called DN in CalledPartyID (Par.13). The terminating service to be provided is determined by DN. A few possible services are outlined below.

Foreign Exchange Line. A business customer B, residing in one city, desires a DN of a line in another city [20]. For example, B may live in a suburb of Chicago and wants to advertise a Chicago DN, so that his Chicago clients can reach him by making a local call. In the past, this required the installation of a physical subscriber line from B's premises to the local Chicago exchange that covered the advertised DN.

In AIN, the service can be implemented without a physical line. B's line is attached to his local exchange, where its directory number is DN_1. The number advertised by B is DN_2, which is covered by a Chicago SSP, where a TAT trigger is assigned to that number.

When a client dials DN_2, the SSP encounters the trigger and sends a Termination_Attempt message, with DN_2 in CalledPartyID (Par.13). SCP returns a Forward_Call message, with DN_1, in CalledPartyID, and SSP forwards the call to B's local exchange.

Call Distribution. Suppose that business customer B employs n "work at home" agents [20]. The directory numbers of the agents, whose lines may be attached to different SSPs, are $DN_1, DN_2, \ldots, DN_n$.

The advertised directory number of B's business is DN_A, a number covered by SSP_A, but which does not represent an actual line. In SSP_A, a TAT trigger is assigned to DN_A, and SCP has a list of DNs associated with DN_A.

A call to DN_A encounters the TAT trigger and causes SSP_A to send a Termination_Attempt message, in which CalledPartyID holds DN_A.

The SCP then checks idle/busy status of the agents on the list, by executing status notification transactions (Section 17.6.4) with the SSPs to which the agents are attached, and allocates the call to an idle agent. Suppose that the previous call has been allocated to DN_5. SCP then starts checking the status of DN_6 and continues until it has found an idle DN or has come full circle without finding an idle DN. In the first case, it sends an Analyze_Route, or a Forward_Call, message to SSP_A, depending on whether the line identified by DN is attached to SSP_A. If all DNs are busy, SCP sends a Disconnect, or Send_To_Resource_Final, message.

17.8 ACRONYMS

AC	Authorization code
AC(3)	Area code (three digits)
AIN	Advanced intelligent network
AMA	Automatic message accounting
ANR	Analyze Route message
ASE	Application service element
BCD	Binary coded decimal
BCM	Basic call model
CCIS	Common-Channel Interoffice Signaling
CDP	Customized dialing plan trigger
DISC	Disconnect message
DN	Directory number
DTMF	Dual tone multifrequency
FCD	Feature code dialing trigger
FWC	Forward Call message
GT	Global title

GTA	Global title address (part of GT)
IAM	Initial Address message (in ISUP signaling)
IC	Interexchange carrier
ID	Identifier
IN	Intelligent network
IN-ASE	Intelligent network application service element
INSC	Intelligent network service circuit
ISDN	Integrated Services Digital Network
ISUP	ISDN User Part
ITU	International Telecommunication Union
ITU-T	Telecommunication Standardization Sector of ITU
LATA	Local access and transport area
NN	Nature of number
NP	Numbering plan
NANP	North American numbering plan
PC-SSN	Point code and subsystem number
O-BCM	Originating basic call model
OHD	Off-hook delay trigger
OHI	Off-hook immediate trigger
PBX	Private branch exchange
PFG	Private facility group
RES-CLR	Resource Clear message
SCE	Service creation environment
SCCP	Signaling Connection Control Part
SCP	Service control point
SMS	Service management system
SSN	Subsystem number
SSP	Service switching point
SS7	Signaling System No. 7
STR-F	Send To Resource, Final message
STR-I	Send To Resource, Interaction message
TAT	Termination attempt trigger
T-BCM	Terminating basic call model
TCAP	Transaction Capability Application Part
TT	Translation type
3/6/10D	Three, six, or ten digit trigger

17.9 REFERENCES

1. R. K. Berman and J. H. Brewster, Perspective on the AIN Architecture, *IEEE Commun. Mag.*, **31**, No. 2, Feb. 1992.
2. S. M. Boyles, R. L. Corn, and L. R. Moseley, Common Channel Signaling; The Nexus of an Advanced Communications Network, *IEEE Commun. Mag.*, **28**, No. 7, July 1990.
3. J. J. Garrahan, P. A. Russo, K. Kitami, and R. Kung, Intelligent Network Overview, *IEEE Commun. Mag.*, **31**, No. 3, Mar. 1993.

4. P. A. Russo, K. Bechard, E. Brooks, R. L. Corn, R. Gove, W. L. Honig, and J. Young, IN Rollout in the United States, *IEEE Commun. Mag.*, **31**, No. 3, Mar. 1993.
5. J. M. Duran and J. Visser, International Standards for Intelligent Networks, *IEEE Commun. Mag.*, **31**, No. 2, Feb. 1992.
6. *General Aspects of the Intelligent Network Application Protocol*, Rec. Q.1208, ITU-T, Geneva, 1993.
7. *Introduction to Intelligent Network Capability Set 1*, Rec. Q.1211, ITU-T, Geneva, 1993.
8. *Interface Recommendation for Intelligent Network CS-1*, Rec. Q.1218, ITU-T, Geneva, 1993.
9. *Advanced Intelligent Network (AIN) Release 1 Switching Systems Generic Requirements*, TA-NWT-001123, Bellcore, Piscataway, NJ, 1991.
10. *Advanced Intelligent Network (AIN) Release 1 Switch—Service Control Point (SCP)/ Adjunct Application Protocol Interface Generic Requirements*, TA-NWT-001126, Bellcore, Piscataway, NJ, 1991.
11. *Ameritech Service Switching Point Functional Specification*, AM-TR-OAT-000042, Ameritech, Hoffman Estates, IL, 1989.
12. *Ameritech Service Control Point Functional Specification*, AM-TR-OAT-000043, Ameritech, Hoffman Estates, IL, 1989.
13. *Ameritech Service Control Point—Service Switching Point Interface Specification*, AM-TR-OAT-000044, Ameritech, Hoffman Estates, IL, 1989.
14. *BellSouth AIN Release 0.1 Service Switching Point Functional Specification*, TR-73562, BellSouth Telecommunications Inc., Birmingham, AL, 1991.
15. *BellSouth AIN Release 0.1 Service Switching Point—Service Control Point Application Protocol Interface Specification*, TR-73563, BellSouth Telecommunications Inc., Birmingham, AL, 1991.
16. *Advanced Intelligent Network (AIN) 0.1 Switching System Generic Requirements*, TR-NWT-001284, Bellcore, Piscataway, NJ, 1992. (replaced by GR-1284)
17. *Advanced Intelligent Network (AIN) 0.1 Switch—Service Control Point (SCP) Interface Generic Requirements*, TR-NWT-001285, Bellcore, Piscataway, NJ, 1992. (replaced by GR-1299)
18. *Advanced Intelligent Network (AIN) 0.2 Switching System Generic Requirements*, TR-NWT-001298, Bellcore, Piscataway, NJ, 1993. (replaced by GR-1284)
19. *Advanced Intelligent Network (AIN) 0.2 Switch—Service Control Point (SCP)/Adjunct Interface Generic Requirements*, TR-NWT-001299, Bellcore, Piscataway, NJ, 1993. (replaced by GR-1299)
20. *Advanced Services Platform, Release 0.1B*. Document AT&T 235-190-126. AT&T Bell Laboratories Technical Publications Center, Naperville, IL, 1995.
21. *The International Public Telecommunication Numbering Plan*, E.164, ITU-T, Geneva, 1997.

18

INTELLIGENT NETWORK APPLICATION PART

18.1 INTRODUCTION

The Intelligent Network Application Part (INAP) is the ITU-T standard protocol used for communication between application entities (AEs) in an intelligent network (IN) [1]. INAP is similar to, but more ambitious than, the North American AIN (advanced intelligent network) described in the previous chapter. A regional (European) variation of INAP is specified by ETSI. INAP is a complex subject so, to put INAP into the proper perspective before delving into the details of the protocol, a brief introduction to the ITU-T architectural model for IN is in order.

The ITU-T architectural model for IN, called the *IN conceptual model* (INCM), describes IN services according to four *planes:*

1. Service plane
2. Global functional plane (GFP)
3. Distributed functional plane (DFP)
4. Physical plane

The four planes represent four different ways of looking at INs and services. They are described in descending order of abstraction, with the service plane providing the most abstract view and the physical plane the least abstract view [2].

The INCM is based on a conceptual model of IN processes that separates the functions that control IN services from the functions that control connections. Service control is the function of the *service logic program* (SLP), which typically runs distributed in IN nodes. Connection control is the function of the *basic call process* (BCP), which typically runs in access nodes (local exchanges). Triggers

Signaling in Telecommunication Networks, Second Edition, by John G. van Bosse and Fabrizio U. Devetak

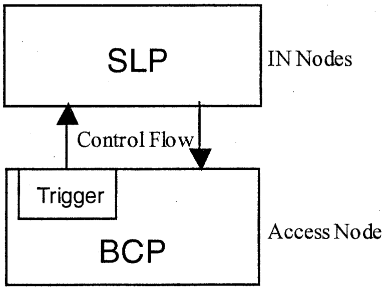


Figure 18.1-1 Conceptual model of IN processes.

in the BCP (called *detection points*—DPs) transfer control to the SLP, from where it is returned to the BCP to complete the connection process (Fig. 18.1-1).

18.1.1 The Four Planes of the IN Conceptual Model

Figure 18.1-2 shows the four IN planes and how they relate to each other.

Service Plane. The service plane defines IN-based services from a user's viewpoint, without any awareness of network components, internal components, or implementation details. Services are built around s*ervice features* [3]. Examples of services are internetwork freephone (IFPH), conference calling (CONF), and universal personal telecommunications (UPT). Examples of service features are call waiting (CW), call forwarding (CF), and call transfer (CT).

Global Functional Plane. The GFP decomposes services into a collection of *service-independent building blocks* (SIBs) that can be chained together to build services and service features [4].

SIBs are service-independent and global. Service independence means that a SIB can be used to build different services. Global means that SIBs are network-wide concepts, unaware of the network elements used to implement them, and of how their functionality may be distributed across those elements. Examples of SIBs are charge, compare, and basic call process.

In this plane a service is realized by linking together multiple SIBs under the control of a subset of the service logic program. The SLP is broken into packages (one per service) called *global service logic* (GSL), which interact with instances of the *basic call process* (BCP), as shown in Fig. 18.1-3. The BCP is a special SIB that controls building and releasing connections and is an abstract monolithic process, unaware of how it may be distributed among functional entities. Instances of the BCP typically run in local exchanges (access nodes) that provide access to users.

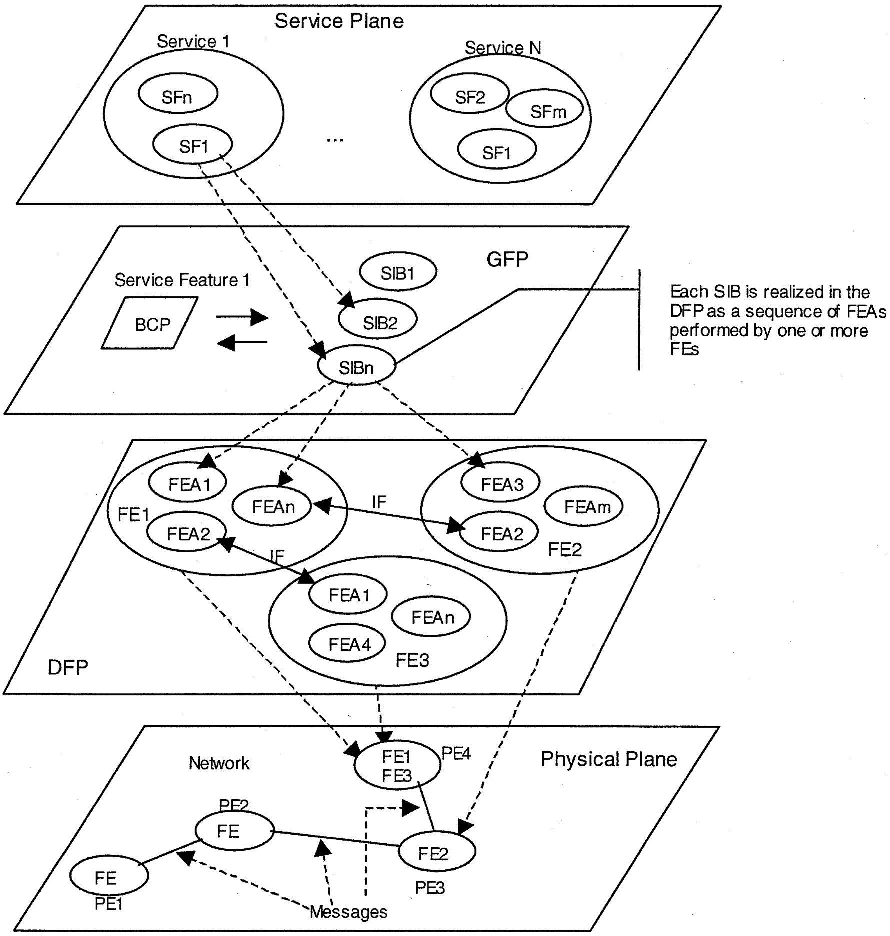


Figure 18.1-2 The four IN planes and their relationships. (From Rec. Q.1201. Reproduced with the kind permission of ITU.)

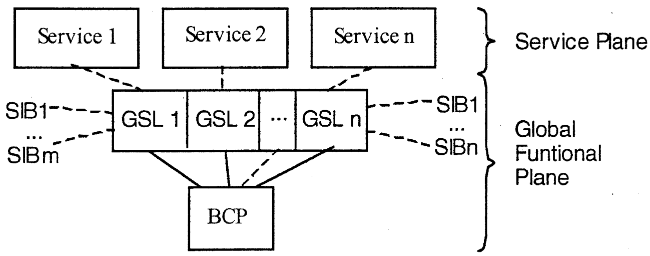


Figure 18.1-3 Representation of SLP in the global functional plane.

The GFP is concerned only with the logic necessary to realize services, not with the functional entities that run these processes or with how they communicate. Where and how SLP and BCP run are progressively detailed in the next two planes.

Distributed Functional Plane. The DFP looks at the IN as a collection of *functional entities* (FEs) that perform *functional entity actions* (FEAs) and exchange *information flows* (IFs) over *relationships*. Functional entities are the logical (functional) elements that perform the functions necessary to implement SIBs.

This view is still an abstraction, as FEs represent functions and do not necessarily map one-to-one to physical entities. Information flows represent the abstract functional view of messages exchanged between FEs, and relationships are the abstract functional view of connections (links) between FEs [5]. Examples of FEs are:

- *Call Control Agent Function* (*CCAF*). It represents user access.
- *Call Control Function* (*CCF*). This non-IN function provides basic call connections (basic call function of a local exchange).
- *Call Unrelated Service Function* (*CUSF*). It manages call unrelated functions with the help of CCF/SSF.
- *Intelligent Access Function* (*IAF*). It provides the logic to connect a non-IN with an IN and resides in the non-IN part.
- *Service Control Function* (*SCF*). It manages the IN service logic (SLP).
- *Service Data Function* (*SDF*). It contains the data managed by the IN service logic.
- *Service Management Function* (*SMF*). It handles management and provisioning of IN services.
- *Service Switching Function* (*SSF*). It contains detection points and provides IN logic to the CCF. SSF represents the local exchange logic needed to support IN service controlled by IN nodes and is typically tightly coupled with CCF.
- *Specialized Resource Function* (*SRF*). It contains resources such as DTMF send/receive and recorded announcements.

In this plane the SLP is subdivided into packages called *distributed service logic* (DSL), with one DSL package for each SIB. As the name implies, a DSL package is distributed among various functional entities.

Physical Plane. The physical plane describes the physical implementation of FEs, mapping FEs into *physical entities* (PEs) that can be built by equipment vendors and purchased by service providers. PEs are connected by data links (the physical implementation of relationships) and exchange INAP signaling messages (physical implementation of IFs) [6]. Examples of PEs are:

- *Service Control Point* (*SCP*). SCF maps into SCP.
- *Service Data Point* (*SDP*). SDF maps into SDP.

- *Intelligent Peripheral* (*IP*). SRF maps into IP.
- *Service Switching Point* (*SSP*). SSF and CCF map into SSP.

The one-to-one mapping of FEs to PEs shown above is not the only one allowed: mapping of more than one FE, such as SCF and SDF or SRF and SSF, into one physical entity are also possible.

In the physical plane the SLP is represented, in a realistic rather than abstract view, as a set of software packages that reside and run in the processors controlling the physical entities.

18.2 CALL MODELS AND TRIGGERS

18.2.1 Call Models

ITU-T specifications for the distributed functional plane use *finite state machines* (FSMs) to specify points in the basic call process (BCP), where interaction with IN service logic is appropriate. States that are relevant to IN activities are referred to as points in call (PICs). Triggers, that is, events at which transfer of control to IN service logic may be in order, are referred to as detection points (DPs). The main FSM model describes CCF activities and is called the *basic call state model* (BCSM). The BCSM is further subdivided into two components:

1. Originating basic call state model (O_BCSM)
2. Terminating basic call state model (T_BCSM)

18.2.2 Triggers (Detection Points)

Transfer of control from the BCSM to a DSL (i.e., from BCP to SLP) is based on detection points. DPs have the potential of transferring control but do not necessarily do so. In order to be activated when specific conditions are met, such as when a certain digit string is received, DPs need to be *armed*. Arming can be *static* (by provisioning) or *dynamic* (by instruction from the SCF). Table 18.2-1 lists examples of DPs from the CS-1 set (see Section 18.3).

TABLE 18.2-1 Detection Points (CS-1)

O_BCSM		T_BCSM	
DP #	Name	DP #	Name
1	Origination_Attempt_Authorized	12	Termination_Attempt_Authorized
5	O_Called_Party_Busy	13	T_Called_Party_Busy
6	O_No_Answer	14	T_No_Answer
7	O_Answer	15	T_Answer
8	O_Mid_Call	16	T_Mid_Call
9	O_Disconnect	17	T_Disconnect
10	O_Abandon	18	T_Abandon

18.3 CAPABILITY SETS

ITU-T recommendations take a phased approach to defining IN-based services. The approach is based on the concept of *capability sets* (CSs). Capability sets group services and service features into sets that are defined in a progression from basic to elaborate, in order to support the evolution of the telephone network, that is, the replacement of older equipment and the introduction of more complex IN-capable hardware and software. ITU-T envisions eight capability sets: CS-1 through CS-8. At the time of this writing, CS-1 through CS-4 have been defined. All CSs are subsets of the total universe of IN-based services. A progressively more robust set of internetworking functionality is required for higher-numbered, compared to lower-numbered, capability sets.

CS-1. This set covers telecommunications services that can be provided in switching environments with legacy exchanges that are not fully equipped for IN services [7–10].

CS-1 services are supported in the PSTN, ISDN, and PLMN and involve the following network interfaces: analog lines, ISDN lines (basic and primary rate), channel-associated signaling trunks, and SS7 trunks.

Service features in this set are *single ended* and have a *single point of control*. That means that a CS-1 service applies to only one party in a call and is not influenced by any other party who may be involved in the same call. Multiple parties in the same call may, however, use the service independently of each other.

Services require interaction only between FEs located in the same network, with the exception of the SCF–SDF interface, which can span across two networks.

The following functional entities are involved in CS-1: CCAF, CCF, SSF, SCF, SDF, and SRF.

Examples of services are listed in Table 18.3-1.

TABLE 18.3-1 CS-1 Services

Service	Acronym
Abbreviated dialing	ABD
Call forwarding	CF
Completion of call to busy subscriber	CCBS[a]
Credit card calling	CCC
Follow-me diversion	FMD
Freephone	FPH
Malicious call identification	MCI
Mass calling	MAS
Selective call forward on busy/no answer	SCF
Televoting	VOT
Terminating call screening	TCS
Universal personal telecommunications	UPT[b]

[a]Only partially supported, due to the single ended, single point of control constraints.
[b]Mainly Personal Number for wireline. Enhanced with wireless in CS-2.

CS-2. This set expands and enhances the telecommunication services covered by CS-1 and introduces two new categories: *service management* and *service creation* services, while still maintaining the "single ended, single point of control" approach [11–15]. To support the new services, an additional FSM is introduced: the *basic call unrelated process* (BCUP), which models FE interactions that are not directly related to setting up a call, such as activities necessary for authentication.

Some of the enhancements in telecommunication services are:

- Terminal mobility (wireless) support
- Call party handling services to allow multiparty control and midcall interaction by users
- Out-of-channel user interaction (e.g., message-waiting indicator and user authentication)

CS-2 services are a superset of CS-1 (see Table 18.3-2 for examples of the additional services). They are built around an extensive set of service features such as user authentication (UAUT), handover (HOV), radio paging (RPAG), and call retrieve (CRET).

CS-2 services are supported by the same interface types as CS-1 but require additional FEs: SMF, IAF, and CUSF.

CS-3. This set broadens the scope of CS-1 and CS-2 by allowing multiple points of control and adding network services for service providers to the purely subscriber services of CS-1 and CS-2 [16,17]. Other enhancements introduced with CS-3 are support of ISDN supplementary services and of number portability.

CS-3 services are a superset of CS-2 and require the same FEs and interface types as CS-1 and CS-2.

The CS-3 recommendations do not list services per se, only service features grouped into broad categories. The most significant (additional to CS-2) are listed in Table 18.3-3.

TABLE 18.3-2 CS-2 Telecommunication Services (Additional to CS-1)

Service	Acronym
Call hold	HOLD
Call transfer	CT
Call waiting	CW
Completion of call to busy subscriber	CCBS[a]
Conference calling	CONF
Hot line	HOT
Internetwork freephone	IFPH
Internetwork televoting	IVOT
Message store and forward	MSF
Universal personal telecommunications	UPT[a]

[a]Adds wireless mobility services to the CS-1 service.

TABLE 18.3-3 Examples of CS-3 Service Features (Additional to CS-2)

	Category	Service Feature	Acronym
User	Basic features	Carrier selection handling CCBS support	CSHND CCBSS
	IN-IP network interworking	Request-to-call CSN	RQTCC
	Number portability	Service provider portability for geographical numbers	SPPGN
	Mobility support	UPT registration with smart card	SCREG
Network	Basic features	Multiple points of control	MPCTR
	Mobility support	Dynamic trigger detection point (TDP) activation/deactivation	DTDPA
	Number portability	Network routing number trigger	NRNTR

TABLE 18.3-4 Examples of CS-4 Service Features (Additional to CS-3)

	Category	Service Feature	Acronym
User	Basic features	Support CUG capability Support for local advertising toward mobile subscribers (IMT2000)	SCUGC SLAMS
	IN-IP internetworking	Internet call waiting Request to call IP	ICWTG RQTCI
Network	Basic features	Call party handling signaling relationship Inter-SCP network capability negotiation	CPHSR ISNCN
	IN-IP interworking	Support for IP addressing	SFIPA
	Cooperation between networks	Network request for temporary connection Private network request to initiate call	NRTC NRIC

CS-4. This set enhances the CS-3 capabilities with support of IP connectivity (Chapter 20) and of supplementary services [18,19]. CS-4 services are a superset of CS-3 and examples of additional features can be found in Table 18.3-4.

18.4 INAP SIGNALING

18.4.1 Overview

The INAP protocol provides the signaling functionality required by the IN application to realize services hosted in distributed (remote from each other) physical elements [20].

The communication function of a functional element contained in a PE consists of one or more application entities (AEs), each comprised of several application service elements (ASEs). The relationship between AEs and ASEs is discussed in Section 14.2. INAP messages are the physical realization of information flows between FEs.

TABLE 18.4-1 CS-1 Operations

CS-1 Operations	Relationship
AddEntry	SCF → SDF
ApplyCharging	SCF → SSF
CallInformationReport	SSF → SCF
CallInformationRequest	SCF → SSF
Connect	SCF → SSF
DirectoryBind	SCF → SDF
DirectoryUnbind	SCF → SDF
FurnishChargingInformation	SCF → SSF
InitialDP	SSF → SCF
InitiateCallAttempt	SCF → SSF
OAnswer	SSF → SCF
OCalledPartyBusy	SSF → SCF
ODisconnect	SSF → SCF
PlayAnnouncement	SCF → SRF
ReleaseCall	SCF → SSF
RemoveEntry	SCF → SDF
SendChargingInformation	SCF → SSF
TAnswer	SSF → SCF
TBusy	SSF → SCF
TDisconnect	SSF → SCF

The IN application uses operations to realize its services (Chapter 14) [21], with each ASE responsible for one or more operations. INAP AEs send requests to invoke operations and respond to requests with *return results* or with the invocation of other operations. Return results are returned for operations that request data from the recipient of the request, such as AddEntry, DirectoryBind, and RemoveEntry (Table 18.4-1); invokes are returned when a request asks for instructions on what to do next. When operations fail or cannot be processed, AEs send back *return error* or *reject* notifications.

Invokes, results, and error and reject notifications are transported in TCAP/SCCP or ISDN messages (see Chapters 16 and 10, respectively). INAP ASEs shape those messages via INAP-TCAP or INAP-ISDN primitives, which carry the operation parameters needed by INAP services and the addressing information needed for delivery of the messages.

Each message type contains the appropriate parameters from the operation template: all types contain the Operation code (which identifies the operation), invokes carry the ARGUMENT, return results contain the RESULT, and return error notifications contain the ERRORS parameter. Operation parameters are based on the ROSE protocol, which is embedded in the Transaction sublayer of TCAP (Chapter 16), or in ISDN call-control messages [22,23]. More on ROSE and Operation parameters can be found in Section 14.1 and in [24–27].

Figure 18.4-3, described in Section 18.4.6, shows how an operation invoke is translated into a TCAP message and how the process is reversed at reception.

INAP signaling is an Application layer (L7) protocol [28] and most commonly uses TCAP and SCCP as transport mechanisms. The other allowed transport, the

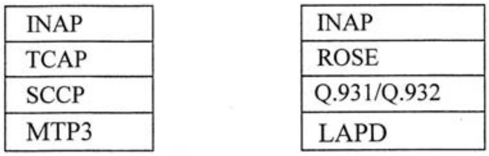


Figure 18.4-1 INAP protocol stacks.

ISDN protocol, may be used between an IP and a SCP (relayed via SSP). Figure 18.4-1 shows how INAP fits in the SS7 and the DSS1 protocol stacks.

ITU-T recommendations describe INAP signaling in terms of three areas:

1. Rules for single and multiple interactions between PEs
2. Operations
3. Procedures (i.e., actions taken after receiving a request for an operation)

A physical entity may have a single interaction with another PE in relation to a given call, in which case the rules controlling the execution of operations are managed by the *single association control function* (SACF)—see Fig. 18.4-2. A PE may also have multiple linked interactions with other PEs in relation to a given call, in which case the rules are managed by the *multiple association control function* (MACF). A typical example of the latter is a SCP interacting with a SSP, a SDP, and an IP on behalf of the same call. The MACF works in concert with the

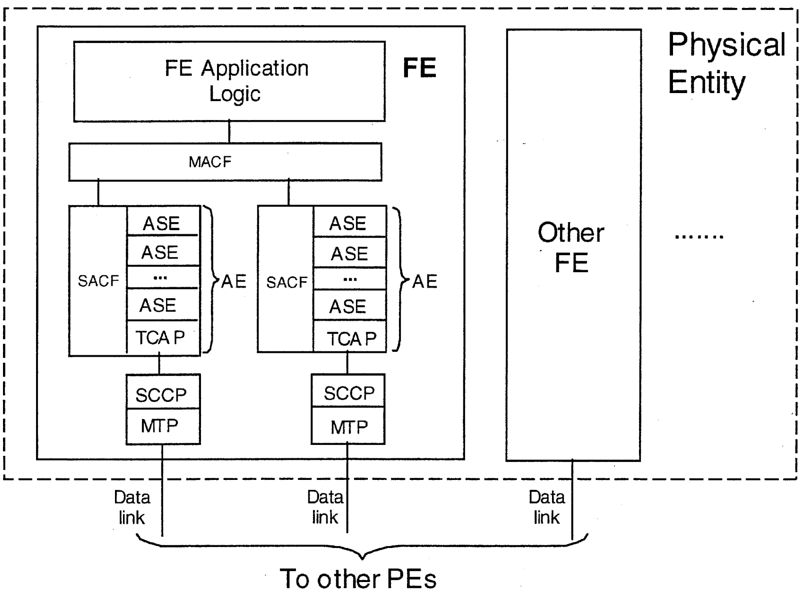


Figure 18.4-2 INAP architecture in a physical entity.

SACF. MACF and SACF are a set of rules that apply to the INAP protocol; for more details we refer the reader to the Q.12x8 set of Recommendations.

INAP provides for backward compatibility with future changes to the protocol by taking advantage of two recent improvements in protocol design: *extensions* and *context negotiation*. Minor additions to the protocol are handled via the extension mechanism, which consists of embedding in message formats optional fields that are left open for future additions, and of defining rules that specify behavior (by the current version) when messages containing the new fields are received. Major additions to the protocol are handled via the application context negotiation mechanism of TCAP, described in Chapter 16.

An example of the INAP protocol architecture in a physical entity is shown in Fig. 18.4-2. The figure shows a case with multiple interactions. For single interaction the MACF is absent or bypassed.

18.4.2 INAP, TCAP, and SCCP

INAP, as a TC-user, uses the structured dialog facility, dialog-handling primitives, and component-handling primitives to communicate with TCAP (Chapter 16). Unstructured dialog is not used [29–33]. Primitives are exchanged under the control of the SACF (Fig. 18.4-2).

An INAP AE conveys the address of the destination AE to TCAP in a dialog-handling TC primitive such as TC-UNI or TC-BEGIN (Section 16.1.3). As explained in Section 14.2.4, the address takes the form of a point code (PC), which identifies the PE, and a subsystem number (SSN), which identifies the AE within the PE. The Operation code parameter identifies the ASE within the AE, since an operation is unique within an AE.

18.4.3 INAP Operations

The following paragraphs show examples of operations for various types of functional element interfaces standardized by ITU-T. Since INAP messages are exchanged between physical entities, the actual messages exchanged can vary with the implementation, depending on which FEs are located in each PE.

ITU-T recommendations define operations using the ASN.1 notation [34–36] and define procedures using state diagrams (SDL). Due to the complexity of the requirements and of the notations, we refer the reader to the Q.12 $\times$ 8 series of recommendations for the detailed formats.

INAP Operations for CS-1. Examples of CS-1 operations are listed in Table 18.4-1 [37,38]. INAP signaling is used for the following interfaces: SCF–SDF, SCF–SSF, and SCF–SRF.

INAP Operations for CS-2. A selection of additional operations is listed in Table 18.4-2 [39,40]. The following additional interfaces are supported: SDF–SDF, SCF–SCF, and SCF–CUSF.

TABLE 18.4-2 CS-2 Operations

CS-2 Operations	Relationships
AssociationRelease Requested	CUSF → SCF
AuthorizeTermination	SSF → SSF
	SRF → SSF
ComponentReceived	CUSF → SCF
DisconnectLeg	SCF → SSF
DSABind	SDF → SDF
Execute	SCF → SDF
FacilitySelectedAndAvailable	SSF → SCF
HandlingInformationRequest	SCF → SCF
HandlingInformationResult	SCF → SCF
MergeCallSegments	SCF → SSF
MoveCallSegments	SCF → SSF
OriginationAttempt	SSF → SCF
SendComponent	SCF → CUSF
TerminationAttempt	SSF → SCF

INAP Operations for CS-3. Additional operations are listed in Table 18.4-3 [41–47]. The same FE interfaces are supported as for CS-2.

INAP Operations for CS-4. Additional operations are listed in Table 18.4-4 [48–54]. The same FE interfaces are supported as for CS-2.

TABLE 18.4-3 CS-3 Operations

CS-3 Operations	Relationship
ChainedRunUserScript	SCF → SCF
ConnectAssociation	SCF → CUSF
EventReportBCUSM	CUSF → SCF
InitialAssociationDP	CUSF → SCF
RequestReport BCUSMEvent	SCF → CUSF
RunUserScript	SCF → SCF
SendSTUI	SCF → CUSF
TrafficFlowControl	SCF → SCF
	SDF → SDF
	SDF → SCF

TABLE 18.4-4 CS-4 Operations

CS-4 Operations	Relationship
Announcement CompletionReport	SCF → SCF
CallFiltering	SCF → SSF
InitiateCallRequest	SCF → SCF
MonitorRouteReport	SSF → SCF
MonitorRouteRequest	SCF → SSF
ProvideAnnouncement Request	SCF → SCF

TABLE 18.4-5 Errors for CS-1, CS-2, CS-3, and CS-4

Error Type	Error	CS
O.R.	ExecutionError	CS-2 thru CS-4
E.R.	Expiration of TSRF	CS-1 thru CS-4
E.R.	Expiration of TSSF	CS-1 thru CS-4
O.R.	IN-ServiceError (Service)	CS-1 thru CS-4
O.R.	MissingParameter	CS-1 thru CS-4
O.R.	ParameterOutOfRange	CS-1 thru CS-4
O.R.	Unexpected ComponentSequence	CS-1 thru CS-4
O.R.	UnavailableResource	CS-1 thru CS-4

Note: O.R., operation-related; E.R., entity-related.

18.4.4 Errors

Two types of error messages are used by INAP: operation-related errors, which are sent in response to an operation request, and functional-entity errors, which are sent for error conditions not related to operations. Table 18.4-5 shows some of the possible INAP errors for CS-1 through CS-4.

18.4.5 Parameters

Due to the complexity of the protocol, we only show a small example of the parameters associated with INAP operations. The parameters in Table 18.4-6 (all part of the Argument parameter field in the operation template) are for the InitialDP operation. InitialDP is sent from a SSF to a SCF after a detection point event triggers a request to the service logic for instructions. All the listed parameters are optional.

18.4.6 Example of INAP Signaling Sequence

INAP message types (e.g., Begin, Continue) are essentially the same as those for AIN (Chapter 16). The same holds for their components (Invoke, Return Result, etc.). Figure 18.4-3 shows an example of a service being realized by the interaction

TABLE 18.4-6 Examples of Parameters for the InitialDP Operation

BearerCapability	LocationNumber
CalledPartyNumber	MiscCallInfo
CallingPartysCategory	OriginalCalledPartyID
CallingPartyNumber	RedirectionInformation
CallingParty Subaddress	RedirectingPartyID
EventTypeBCSM	ServiceInteraction Indicators
Forward CallIndicators	ServiceKey
HighLayer Compatibility	ServiceProfileIdentifier
IPSSPCapabilities	TerminalType
IPAvailable	TriggerType

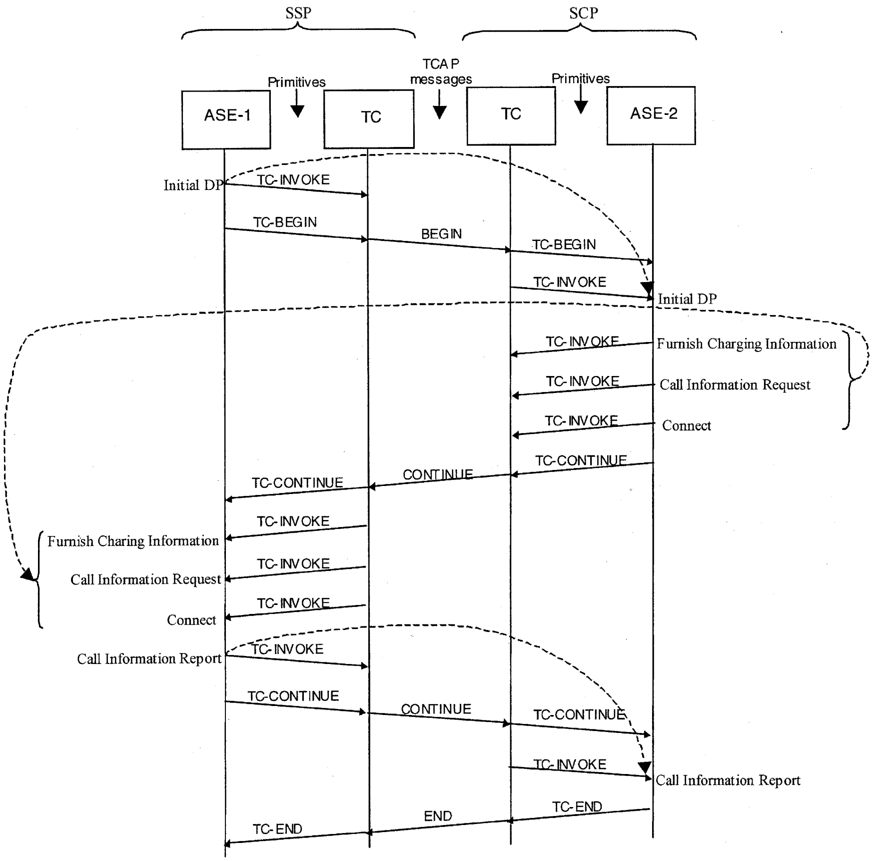


Figure 18.4-3 Example of INAP signaling exchange.

of a SSP, containing a SSF/CCF, with a SCP containing a SCF. The SSP and the SCP are connected by SS7 signaling links. Only the TC layers are shown, and for an explanation of TC dialogs and messages we refer the reader to Chapter 16. The service in this example is freephone in a case where the SCP calculates charges. The SSP detects a triggering event (DP) based on the dialed digits and decides to invoke an InitialDP operation from the SCP. Upon receipt of the operation request, the SCP responds with the invocation of three operations, all transported in the same TC message:

1. FurnishChargingInformation, which specifies, in one of the parameters, the type of charging information to be collected (in this case the called party is charged)
2. Connect, which instructs the SSP to establish the connection between the calling and the called party
3. CallInformationRequest, which instructs the SSP to send information to the SCP at the end of the call so it can compute the charges

At the end of the call the SSP invokes the CallInformationReport operation, populating the Argument parameter field with the charge information that the SCP had requested. The final End message does not contain any operation and closes the TC dialog.

18.5 ETSI INAP

The European standards agency ETSI has issued a suite of standards for CS-1, CS-2, CS-3, and CS-4 [55–61].

ETSI INAP standards for CS-1 follow closely the ITU-T specifications, with only a few operations not included.

The ETSI CS-2 set has most (although not all) of the additional CS-2 operations of the ITU-T set and also includes some ITU-T CS-3 operations.

The ETSI CS-3 and CS-4 sets are subsets of the corresponding ITU-T sets.

18.6 ACRONYMS

AE	Application entity
AIN	Advanced intelligent network
APDU	Application protocol data unit
ASE	Application service element
ASN.1	Abstract Syntax Notation One
BCP	Basic call process
BCSM	Basic call state model
BCUP	Basic call unrelated process
CCAF	Call control agent function
CCF	Call control function
CF	Call forwarding
CONF	Conference calling
CRET	Call retrieve
CS-*n*	Capability set *n*
CT	Call transfer
CUSF	Call unrelated service function
CW	Call waiting
DFP	Distributed functional plane
DP	Detection point
DSL	Distributed service logic
DSS1	Digital subscriber Signaling System No. 1
ETSI	European Telecommunications Standards Institute
FE	Functional entity

FEA	Functional entity action
FSM	Finite state machine
GFP	Global functional plane
GSL	Global service logic
HOV	Handover
IAF	Intelligent access function
IF	Information flow
IFPH	Internetwork freephone
IN	Intelligent network
INAP	Intelligent Network Application Part
INCM	IN conceptual model
IP	Intelligent peripheral, Internet Protocol
ISDN	Integrated Services Digital Network
ISUP	ISDN User Part
ITU	International Telecommunication Union
ITU-T	Telecommunication Standardization Sector of ITU
MACF	Multiple association control function
MTP	Message Transfer Part
O_BCSM	Originating basic state call model
PC	Point code
PE	Physical entity
PIC	Point in call
PLMN	Public land mobile network
PSTN	Public switched telephone network
ROSE	Remote Operations Service Element
RPAG	Radio paging
SACF	Single association control function
SCCP	Signaling Connection Control Part
SCF	Service control function
SCP	Service control point
SDF	Service data function
SDL	Specification and description language
SDP	Service data point
SIB	Service-independent building block
SLP	Service logic program
SMF	Service management function
SRF	Specialized resource function
SSF	Service switching function
SSN	Subsystem number

SSP	Service switching point
T_BCSN	Terminating basic state call model
TC	Transaction capabilities
TCAP	Transaction Capabilities Application Part
UAUT	User authentication
UPT	Universal personal telecommunications

18.7 REFERENCES

1. Igor Faynberg, Marc P. Kaplan, Lawrence Gabuzda, and Nitin Shah, *The Intelligent Network Standards: Their Application to Services*, McGraw-Hill, New York, 1996.
2. *Principles of Intelligent Network Architecture*, Rec. Q.1201, ITU-T, Geneva, Oct. 1992.
3. *Intelligent Network—Service Plane Architecture*, Rec. Q.1202, ITU-T, Geneva, Sep. 1997.
4. *Intelligent Network—Global Functional Plane Architecture*, Rec. Q.1203, ITU-T, Geneva, Sep. 1997.
5. *Intelligent Network Distributed Functional Plane Architecture*, Rec. Q.1204, ITU-T, Geneva, Mar. 1993.
6. *Intelligent Network Physical Plane Architecture*, Rec. Q.1205, ITU-T, Geneva, Mar. 1993.
7. *Introduction to Intelligent Network Capability Set 1*, Rec. Q.1211, ITU-T, Geneva, Mar. 1993.
8. *Global Functional Plane for Intelligent Network CS-1*, Rec. Q.1213, ITU-T, Geneva, Oct. 1995.
9. *Distributed Functional Plane for Intelligent Network CS-1*, Rec. Q.1214, ITU-T, Geneva, Oct. 1995.
10. *Physical Plane for Intelligent Network CS-1*, Rec. Q.1215, ITU-T, Geneva, Oct. 1995.
11. *Introduction to Intelligent Network Capability Set 2*, Rec. Q.1221, ITU-T, Geneva, Sep. 1997.
12. *Service Plane for Intelligent Network Capability Set 2*, Rec. Q.1222, ITU-T, Geneva, Sep. 1997.
13. *Global Functional Plane for Intelligent Network Capability Set 2*, Rec. Q.1223, ITU-T, Geneva, Sep. 1997.
14. *Distributed Functional Plane for Intelligent Network Capability Set 2*, Rec. Q.1224, ITU-T, Geneva, Sep. 1997.
15. *Physical Plane for Intelligent Network Capability Set 2*, Rec. Q.1225, ITU-T, Geneva, Sep. 1997.
16. *Introduction to Intelligent Network Capability Set 3*, Rec. Q.1231, ITU-T, Geneva, Dec. 1999.
17. *Intelligent Network Capability Set 3—Management Information Model Requirements and Methodology*, Rec. Q.1236, ITU-T, Geneva, Feb. 1999.
18. *Introduction to Intelligent Network Capability Set 4*, Rec. Q.1241, ITU-T, Geneva, July 2001

19. *Distributed Functional Plane for Intelligent Network Capability Set 4*, Rec. Q.1244, ITU-T, Geneva, July 2001.
20. *General Aspects of the Intelligent Network Application Protocol*, Rec. Q.1208, ITU-T, Geneva, Sep. 1997.
21. *Remote Operations: Concept, Model and Notation*, Rec. X.880, ITU-T, Geneva, 1994.
22. *ISDN User-Network Interface Layer 3 Specification for Basic Call Control*, Rec. Q.931, ITU-T, Geneva, May 1998.
23. *Digital Subscriber Signalling System No.1—Generic Procedures for the Control of ISDN Supplementary Services*, Rec. Q.932, ITU-T, Geneva, May 1998.
24. *Remote Operations: Model, Notation and Service Definition*, Rec. X.219, ITU-T, Geneva, 1988.
25. *Remote Operations: Protocol Specification*, Rec. X.229, ITU-T, Geneva, Nov. 1988.
26. *Remote Operations: OSI Realizations—Remote Operations Service Element (ROSE) Service Definition*, Rec. X.881, ITU-T, Geneva, 1994.
27. *Remote Operations: OSI Realizations—Remote Operations Service Element (ROSE) Protocol Specification*, Rec. X.882, ITU-T, Geneva, 1994.
28. *Information Technology—Open System Interconnection—Basic Reference Model: The Basic Model*, Rec. X.200, ITU-T, Geneva, 1994.
29. *Functional Description of Transaction Capabilities*, Rec. Q.771, ITU-T, Geneva, 1997.
30. *Transaction Capabilities Information Element Definition*, Rec. Q.772, ITU-T, Geneva, 1997.
31. *Transaction Capabilities Formats and Encoding*, Rec. Q.773, ITU-T, Geneva, 1997.
32. *Transaction Capabilities Procedures*, Rec. Q.774, ITU-T, Geneva, 1997.
33. *Guidelines for Using Transaction Capabilities*, Rec. Q.775, ITU-T, Geneva, 1997.
34. *Information Technology—Abstract Syntax Notation One (ASN.1): Specification of Basic Notation*, Rec. X.680, ITU-T, Geneva, 2002.
35. *Information Technology—Abstract Syntax Notation One (ASN.1): Information Object Specification*, Rec. X.681, ITU-T, Geneva, 2002.
36. *Information Technology—ASN.1 Encoding Rules: Specification of Basic Encoding Rules (BER), Canonical Encoding Rules (CER) and Distinguished Encoding Rules (DER)*, Rec. X.690, ITU-T, Geneva, 2002.
37. *Interface Recommendation for Intelligent Network CS-1*, Rec. Q.1218, ITU-T, Geneva, Oct. 1995.
38. *Intelligent Network User's Guide for Capability Set 1*, Rec. Q.1219, ITU-T, Geneva, Apr. 1994.
39. *Interface Recommendation for Intelligent Network Capability Set 2*, Rec. Q.1228, ITU-T, Geneva, Sep. 1997.
40. *Intelligent Network User's Guide for Capability Set 2*, Rec. Q.1229, ITU-T, Geneva, Mar. 1999.
41. *Interface Recommendation for Intelligent Network Capability Set 3: Common Aspects*, Rec. Q.1238.1, ITU-T, Geneva, June 2000.
42. *Interface Recommendation for Intelligent Network Capability Set 3: SCF–SSF Interface*, Rec. Q.1238.2, ITU-T, Geneva, June 2000.
43. *Interface Recommendation for Intelligent Network Capability Set 3: SCF–SRF Interface*, Rec. Q.1238.3, ITU-T, Geneva, June 2000.

44. *Interface Recommendation for Intelligent Network Capability Set 3: SCF–SDF Interface*, Rec. Q.1238.4, ITU-T, Geneva, June 2000.
45. *Interface Recommendation for Intelligent Network Capability Set 3: SDF–SDF Interface*, Rec. Q.1238.5, ITU-T, Geneva, June 2000.
46. *Interface Recommendation for Intelligent Network Capability Set 3: SCF–SCF Interface*, Rec. Q.1238.6, ITU-T, Geneva, June 2000.
47. *Interface Recommendation for Intelligent Network Capability Set 3: SCF–CUSF Interface*, Rec. Q.1238.7, ITU-T, Geneva, June 2000.
48. *Interface Recommendation for Intelligent Network Capability Set 4: Common Aspects*, Rec. Q.1248.1, ITU-T, Geneva, July 2001.
49. *Interface Recommendation for Intelligent Network Capability Set 4: SCF–SSF Interface*, Rec. Q.1248.2, ITU-T, Geneva, July 2001.
50. *Interface Recommendation for Intelligent Network Capability Set 4: SCF–SRF Interface*, Rec. Q.1248.3, ITU-T, Geneva, July 2001.
51. *Interface Recommendation for Intelligent Network Capability Set 4: SCF–SDF Interface*, Rec. Q.1248.4, ITU-T, Geneva, July 2001.
52. *Interface Recommendation for Intelligent Network Capability Set 4: SDF–SDF Interface*, Rec. Q.1248.5, ITU-T, Geneva, July 2001.
53. *Interface Recommendation for Intelligent Network Capability Set 4: SCF–SCF Interface*, Rec. Q.1248.6, ITU-T, Geneva, July 2001.
54. *Interface Recommendation for Intelligent Network Capability Set 4: SCF–CUSF Interface*, Rec. Q.1248.7, ITU-T, Geneva, July 2001.
55. *Intelligent Network (IN); Intelligent Network Capability Set 1 (CS1); Core Intelligent Network Application Protocol (INAP); Part 1: Protocol Specification*, ETS 300 374-1, ETSI, Sophia Antipolis, France, Sep. 1994.
56. *Intelligent Network (IN); Intelligent Network Application Protocol (INAP); Capability Set 2 (CS2); Part 1: Protocol Specification*, EN 301 140-1 V1.3.4, ETSI, Sophia Antipolis, France, June 1999.
57. *Intelligent Network (IN); Intelligent Network Capability Set 3 (CS3); Intelligent Network Application Protocol (INAP); Protocol Specification; Part 2: SCF–SSF Interface*, EN 301 931-2 V1.1.2, ETSI, Sophia Antipolis, France, Sep. 2001.
58. *Intelligent Network (IN); Intelligent Network Capability Set 3 (CS3); Intelligent Network Application Protocol (INAP); Protocol Specification; Part 3: SCF–SRF Interface*, EN 301 931-3 VI.1.2, ETSI, Sophia Antipolis, France, Sep. 2001.
59. *Intelligent Network (IN); Intelligent Network Capability Set 3 (CS3); Intelligent Network Application Protocol (INAP); Protocol Specification; Part 4: SDLs for SCF–SSF Interface*, EN 301 931-4 V1.1.2, ETSI, Sophia Antipolis, France, Sep. 2001.
60. *Intelligent Network (IN); Intelligent Network Capability Set 4 (CS4); Intelligent Network Application Protocol (INAP); Protocol Specification; Part 1: Common Aspects*, EN 302 039-1 V1.1.1, ETSI, Sophia Antipolis, France, Nov. 2001.
61. *Intelligent Network (IN); Intelligent Network Capability Set 4 (CS4); Intelligent Network Application Protocol (INAP); Protocol Specification; Part 2: Service Switching Function–Switching Control Function (SSF–SCF) Interface*, EN 302 039-2 V1.1.2, ETSI, Sophia Antipolis, France, Nov. 2001.

19

MOBILE APPLICATION PART

Chapter 12 has described the signaling between a mobile station (MS) and a mobile switching center (MSC) in a cellular (public land) mobile network (CMN, PLMN). The present chapter describes the Mobile Application Part (MAP) of Signaling System No. 7 (SS7). MAP defines a number of remote operations (transactions) that support mobile telecommunications.

Public switched telecommunication networks (PSTNs) are *fixed* networks. A subscriber line is fixed (attached) to a local exchange, which has a record for each of its subscribers. The record contains semipermanent and temporary data. Semipermanent data in the record include the subscriber's directory number and a *service profile*. This is a list of features (e.g., call waiting) which the subscriber has requested, and the telecom has agreed to supply, for a monthly or per-use charge.

The temporary data include information on the current subscriber state (busy, idle, denied service) and, during the times that the subscriber is involved in a call, details about the connection. The exchange consults these data when processing the subscriber's calls.

When a MSC is serving a MS, it needs similar data. However, a MS is "mobile." At different points in time, it can be in the service area of different MSCs. It is not practical to store the information on all mobiles at every MSC. Instead, this information is stored in centralized databases. One of the purposes of MAP transactions is to enable the MSCs to obtain information about a MS from these databases.

This chapter describes two versions of MAP. The U.S. version, denoted here as IS-MAP, is defined by the Electronic Industries Alliance (EIA) and the Telecommunications Industry Association (TIA). The international version, used in the Global System for Mobile Communications (GSM), and denoted here as GSM-MAP, is standardized by the International Telecommunication Union (ITU-T) and the European Telecommunications Standards Institute (ETSI).

IS-MAP is discussed in Sections 19.1–19.5. The description of GSM-MAP starts in Section 19.6.

Signaling in Telecommunication Networks, Second Edition, by John G. van Bosse and Fabrizio U. Devetak

19.1 INTRODUCTION TO IS-MAP

This section introduces some general MAP concepts and terms.

19.1.1 Definitions

First, a few definitions [1] are given.

Home Network. This is the CMN selected for mobile communication services by the owner of a MS.

Roaming. A MS can receive service in its home CMN and in other CMNs. When a MS is being served by a CMN other than its home CMN, we say that the MS is roaming. A MS stores the system identification (SID) of its home CMN—see Section 12.3. A MS knows that it is "at home" if the SID received in overhead parameter messages (Section 12.4.2) matches its stored SID. If the SIDs do not match, the MS is roaming.

Visited Network. This is the CMN that is serving the MS when it is roaming.

Home MSC. This is the MSC in the home CMN of the MS to which the PSTN delivers calls to the MS.

Serving MSC. This is the MSC that is currently serving the MS. Depending on the MS location, this can be the home MSC, another MSC in the home CMN, or a MSC in another CMN.

19.1.2 Equipment Entities Involved in IS-MAP Transactions

The MAP-related equipment entities in a cellular mobile network (CMN) are shown in Fig. 19.1-1.

Mobile Switching Centers (MSCs). In this example, the CMN has three MSCs, each of which serves MSs in its part of the CMN service area.

Home Location Register (HLR). This register stores semipermanent and temporary data on mobiles for which the CMN is the *home* network.

Visitor Location Register (VLR). A VLR is associated with one or more MSCs. In Fig. 19.1-1, VLR-P is associated with MSC-A and MSC-B, and VLR-Q is associated with MSC-C. A VLR stores information on roaming mobiles that are currently being served by an associated MSC.

Authentication Center (AUC). The AUC stores the authentication and voice privacy information of mobiles for which CMN is the home network and which have authentication and voice privacy service (Sections 12.6.7 and 12.6.8).

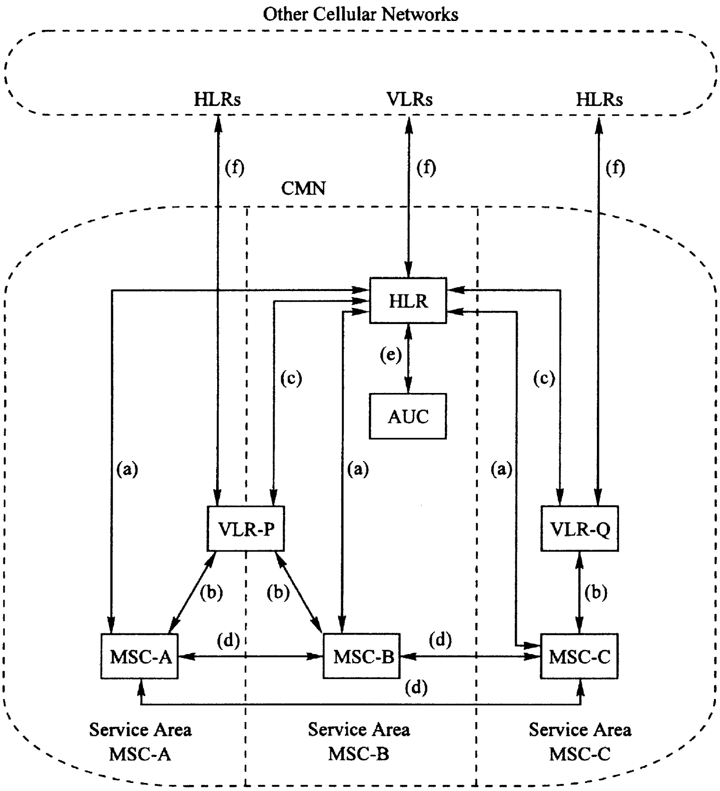


Figure 19.1-1. IS-MAP related equipment entities and interfaces.

In principle, all entities in the figure can be stand-alone units. They are SS7 signaling points and are connected by signaling links to a SS7 signaling network, which transfers the transaction messages. This is the configuration assumed in the material that follows.

Actual equipment configurations can be slightly different. For instance, authentication center AUC, shown here as a stand-alone unit, can be implemented as an integral part of HLR. In this case, AUC is not a signaling point, and the transaction messages between HLR and AUC are internal messages that are passed in primitives. Also, a VLR is sometimes implemented as a part of its associated MSC (e.g., VLR-Q and MSC-C).

19.1.3 Interfaces

The lines in Fig. 19.1-1 indicate the interfaces between equipment units for various transactions. Within a CMN, there are transactions involving a MSC and the HLR (a), a MSC and its associated VLR (b), a VLR and the HLR (c), two adjacent

MSCs (d), and HLR and AUC (e). In addition, there are transactions involving a VLR in one CMN and the HLR of another CMN (f).

19.1.4 Identification of MS, CMN, MSC, and VLR

In the United States, the numbering plan for mobiles is integrated with the PSTN numbering plan (Section 1.2.1). The mobile identification number (MIN) of a MS consists of a three-digit area code (AC), a three-digit exchange code (EC), and a four-digit station code. The MINs in this chapter are the counterparts of MIN1s and MIN2s (binary numbers) in messages on the RF channels between MS and base station (Section 12.4.9), but are formatted as ten BCD digits. When necessary, a MSC converts MIN to MIN1 + MIN2, or vice versa.

The PSTN regards a MSC as a local exchange, which is identified by one or more AC-EC combinations. A MSC identified by, say, 708-371 and 708-456 is the home MSC of mobiles with MINs 708-371-*XXXX* and 708-456-*XXXX*. When a MS is called, the PSTN network routes the call to the home MSC of MS.

U.S. cellular networks are identified by a 15-bit *system identification* (SID—see Section 12.3). Within a CMN, a MSC or VLR is identified by an eight-bit *switch identification* (SWID). The combination SID-SWID is known as the *MSC identification* (MSCID), and uniquely identifies a MSC or VLR in the United States.

19.1.5 Contents of HLR, VLR, and AUC

This section summarizes the most important parameters of the MS records in the various registers.

Home Location Register of CMN. This register has records for all mobiles for which CMN is the home network:

- Mobile identification number (MIN)
- Mobile serial number (MSN)—see Section 12.3
- MS status (qualified or not qualified for service)
- MS service profile
- If MS is currently served by a MSC in its home CMN, the MSCID (and optionally, the point code) of MSC
- If MS is currently served by a MSC in a visited CMN, the MSCID (and optionally, the point code) of the VLR associated with the MSC

All parameters are semipermanent except MSCID, which is temporary.

Visitor Location Register. This register has records for all roaming mobiles that are being served by an associated MSC:

- Mobile identification number (MIN)
- Mobile serial number (MSN)

- MS service profile
- MSCID, and possibly PC, of the associated MSC that currently serves MS

All parameters are temporary. They are entered when a MSC associated with VLR starts serving a roaming MS and are deleted when the service ends.

Authentication Center of CMN. This center has records for all mobiles for which CMN is the home network and that have authentication and voice privacy service:

- MIN
- SSD_A, shared secret data for MS authentication (Section 12.6.7)
- SSD_B, shared secret data for voice privacy (Section 12.6.8)

All parameters are semipermanent.

19.1.6 IS-MAP Operations

Like advanced intelligent network (AIN) operations (Chapter 17), most IS-MAP transactions consist of one operation and involve a query and a response package. Unlike AIN operations, whose response packages contain invokes, IS-MAP response packages contain a return-result or a return-error.

At this time, over 30 IS-MAP operations have been defined. Sections 19.2–19.5 discuss operations for MS registration and authentication, for making calls to and from MS, and for MS handover.

19.1.7 Transfer of TCAP Packages

TCAP uses the connectionless service of the SS7 signaling connection control part (SCCP)—see Section 15.3.

To SCCP, each MSC, HLR, VLR, and AUC is a subsystem at a signaling point and is identified by a point code (PC) and a subsystem number (SSN). In United States cellular networks, the SSN codes that identify the various equipment types have been standardized [1].

SSN	Equipment Type
5	Mobile Application Part (MAP)
6	Home location register (HLR)
7	Visitor location register (VLR)
8	Mobile switching center (MSC)
10	Authentication center (AUC)

SCCP Called Party Addresses. The ASE that initiates a transaction has to determine the SCCP address (CDA) of the called ASE. In transactions between

ASEs in the same cellular network, the initiating ASE provides a PC_SSN (point code and subsystem number) called address.

In transactions involving called ASEs in different CMNs, the called address is a global title (GT), which has to be translated at a signal transfer point.

The examples of the IS-MAP transactions in Sections 19.2–19.4 include information on how the initiating ASE determines the CDA.

19.2 TRANSACTIONS FOR REGISTRATION AND AUTHENTICATION

This section discusses operations that are triggered when a MSC receives a registration or origination message from a MS.

The transactions involve the HLR in the home network of MS, the MSC that received the MS message, and, if MS is roaming, the VLR associated with MSC.

When a MSC of network CMN receives a message from a MS, it analyzes the AC-EC combination of the received MIN. If it recognizes AC-EC as one of the combinations for MINs of mobiles whose home is CMN, it concludes that CMN is the home network of MS; otherwise, MS is roaming.

19.2.1 Registration Notification

This operation is invoked by a MSC and executed by the HLR in the home network of MS. It has several purposes:

- To determine whether MS is qualified to receive service. If yes,
- To transfer the MS service profile to MSC, possibly via its associated VLR, and
- To update HLR information about where the MS is receiving service.

We first consider the case that the mobile and/or the mobile network do not have authentication capability (Section 12.6.7). On receipt of a Registration message from MS, the MSC initiates a "registration notification" transaction [2,3].

In Fig. 19.2-1, the home network of the registering MS is CMN-1. MSC-A and HLR are in the home CMN of MS, and HLR is the source of information about MS. CMN-2 and CMN-3 are other cellular networks. VLR-P and VLR-Q are associated with, respectively, MSC-B and MSC-C.

The TCAP packages with an invoke and a return-result for this operation are denoted by REGNOT and regnot, respectively.

Three cases are considered. In case (*a*), the mobile has registered at MSC-A, which then initiates a registration notification transaction with HLR by sending a REGNOT. In case (*b*), the mobile has registered at MSC-B, which initiates a transaction with VLR-P. We assume that VLR-P has no information about MS and therefore initiates a transaction with HLR. In case (*c*), VLR-P has information about MS and does not need to contact HLR.

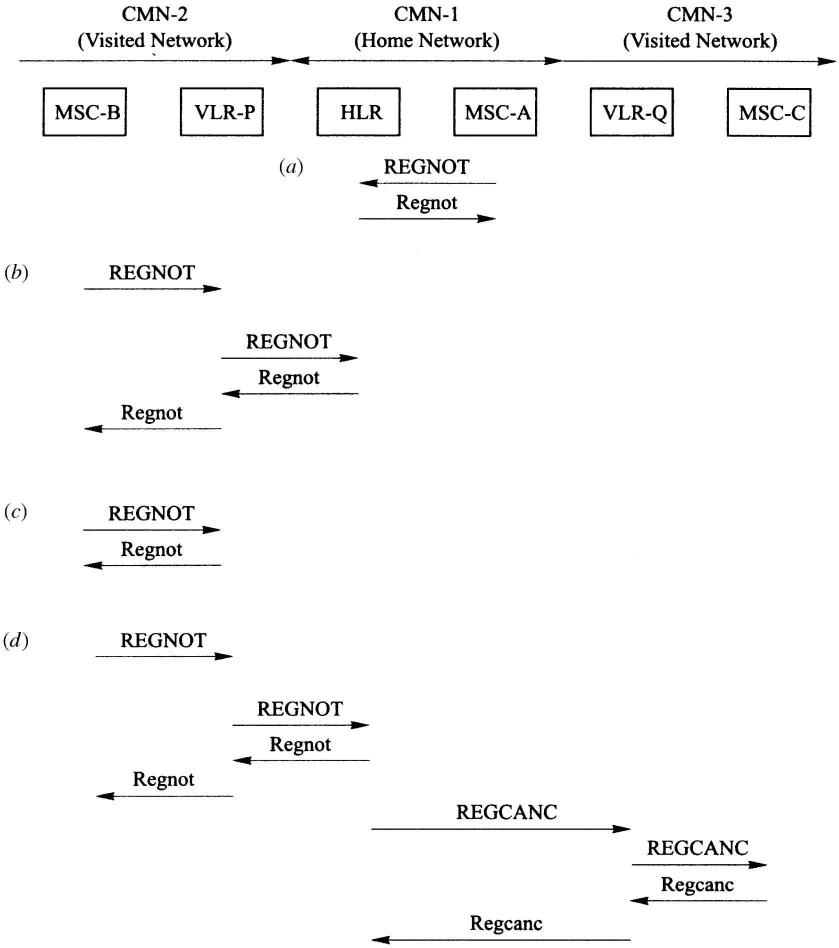


Figure 19.2-1. Registration procedures.

On receipt of REGNOT, HLR first checks whether it has a record for a mobile identified by MIN. If so, it checks whether the mobile is qualified to receive service and—if this is the case—authorizes the invoking MSC to serve the MS for a certain *authorization period*. The regnot return-result includes the authorization period and, if requested in REGNOT, the service profile of MS.

MSC-A or MSC-B and VLR receive the regnot and establish a temporary record with the authorization period and service profile of MS.

While processing REGNOT, HLR also updates its record, by entering the MSCID of MSC-A or VLR-P. The new serving MSC now has the information necessary to process calls for MS, and HLR has updated its information on the whereabouts of MS.

On receipt of regnot, MSC responds to MS, with a Registration Acknowledgment message.

If HLR determines that MS is not qualified to receive service, it indicates this in its regnot, and MSC then sends a Release message to MS.

19.2.2 Expiration of Authorization Period

MSC-A, or MSC-B and VLR-P, periodically examine their records on MS and check whether the authorization period has expired. If this is the case, they discard the record.

On expiration, HLR removes the MSCID, which identifies MSC-A or VLR-P, from its record.

19.2.3 Registration Cancellation

Now assume that MS has registered at MSC-B at a time that the authorization period for MS at the previous serving MSC-C has not yet expired—Fig. 19.2-1(*d*). This is known to HLR, because the MSCID of VLR-Q is in its MS record.

On receipt of the REGNOT, HLR executes the notification operation and also initiates a registration cancellation transaction with VLR-Q, by sending a REGCANC invoke. VLR-Q initiates a cancellation transaction with MSC-C, which then discards its record for MS. On receipt of regcanc, VLR-Q returns a regcanc to HLR and discards its record. When HLR receives the regcanc, it deletes the MSCID of VLR-Q in its MS record.

19.2.4 MS Authentication

Now consider the case that the MS and the MSC have authentication capability (Section 12.6.7), and possibly voice privacy capability (Section 12.6.8). Figure 19.2-2 assumes that CMN-1 is the home network of MS, and that MS intends to register at MSC-B, in network CMN-2.

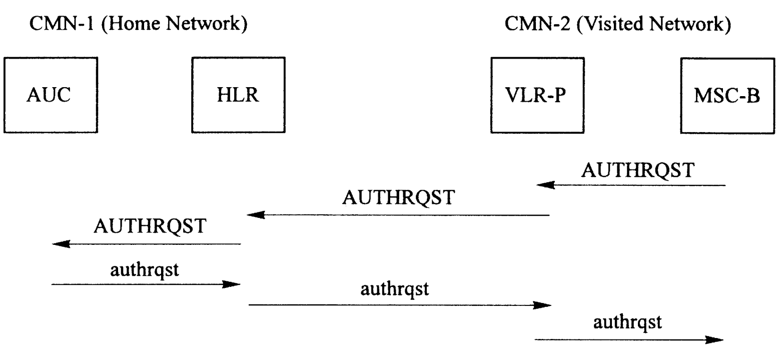


Figure 19.2-2. Authentication of a mobile. (From TSB-51. Reproduced with permission of TIA.)

Before registering, MS executes the CAVE algorithm, using a random number RAND received in an overhead parameter message, its identification and serial numbers (MIN and MSN), and the internally stored shared secret data parameter SSD_A (Section 12.6.7). It includes the authentication result AUTHR in its registration message to MSC-B [4].

MSC-B then starts an authorization-request transaction with VLR-P, with an AUTHRQST invoke. In turn, VLR-P initiates a transaction with HLR, and HLR initiates a transaction with authentication center AUC.

The AUTHRQSTs include AUTHR and all parameters used by MS to calculate AUTHR, except SSD_A. AUC executes CAVE, using the received parameters, and the SSD_A value in its record for MS. If the result matches the received AUTHR, AUC concludes that the mobile stores the correct SSD_A and is therefore authentic. In its authrqst response, AUC indicates whether the mobile is authentic or should be denied access.

This information is passed to MSC-B. If MS is authentic, MSC-B returns a Registration Confirmation MS and starts the registration operation—see Fig. 19.2-1(*b*). If the authrqst indicates that the mobile should not be served, MSC-B returns a release message to MS.

If the mobile is authentic and has voice privacy service, the authrqst includes two voice privacy masks (Section 12.6.8).

The same procedure, but without involving a VLR, applies when MS is served by an MSC in its home network.

19.2.5 Message Contents

This section lists the most important mandatory (M) and optional (O) parameters in the invokes and return-result discussed so far, and in the return-error message, which is returned to the invoker if the requested operation cannot be carried out [2].

In conformity with the EIA/TIA documents, the operation names and parameter tags are denoted as character strings in which the initial letter of each word is in uppercase.

The references Par.1 and so on point to parameter format descriptions in Section 19.5.4. It is helpful to look up the descriptions while reading material that follows.

RegistrationNotification Invoke (REGNOT). See Fig. 19.2-3. The MS is identified by Par.14: MobileIdentificationNumber (MIN) and Par.15: MobileSerialNumber (MSN). The QualificationInformationCode (Par.19) is used to request the validation of the MS and/or the MS service profile from the HLR.

The invoker includes its identity MSCID (Par.16) and, optionally, its PC_SSN (point code + subsystem number) address (Par.18). The responder uses this when sending its return-result or return-error message.

RegistrationNotification Return-Result (regnot). See Fig. 19.2-3. If Par.19 in the invoke has requested validation information, the return-result includes Par.4

Operation		Timer: 6 seconds
RegistrationNotification (REGNOT)		
Invoke Parameters	**M/O**	**Reference**
MobileIdentificationNumber (MIN)	M	Par.14
MobileSerialNumber (MSN)	M	Par.15
MSCID (serving MSC or VLR)	M	Par.16
PC_SSN (serving MSC or VLR)	O	Par.18
QualificationInformationCode	M	Par.19
Return-Result Parameters	**M/O**	**Reference**
AuthorizationDenied	O	Par.3
AuthorizationPeriod	O	Par.4
CallingFeaturesIndicator	O	Par.6
OriginationIndicator	O	Par.17
TerminationRestrictionCode	O	Par.26
Return-Error Information	**M/O**	**Reference**
Error Code	O	Section 19.5.3
FaultyParameter	O	Par.12

Figure 19.2-3. Parameters in registration notification messages. (From IS-41.5-B. Reproduced with permission of TIA.)

(AuthorizationPeriod) or Par.3 (AuthorizationDenied). Authorization is denied when the MS record indicates a delinquent account, or a stolen unit, and so on.

If the registration is accepted and Par.19 in the invoke requests the service profile of the MS, the return-result also includes Par.17 (OriginationIndicator), Par.26 (TerminationRestrictionCode), and Par.6 (CallingFeaturesIndicator).

RegistrationCancellation Invoke (REGCANC). See Fig. 19.2-4. The message identifies the mobile by Par.14 and Par.15 (MIN and MSN).

Operation		Timer: 6 seconds
RegistrationNotification (REGCANC)		
Invoke Parameters	**M/O**	**Reference**
MobileIdentificationNumber (MIN)	M	Par.14
MobileSerialNumber (MSN)	M	Par.15
Return-Result Parameters	**M/O**	**Reference**
None		
Return-Error Information	**M/O**	**Reference**
Error Code	M	Section 19.5.3
FaultyParameter	O	Par.12

Figure 19.2-4. Parameters in registration cancellation messages. (From IS-41.5-B. Reproduced with permission of TIA.)

RegistrationCancellation Return-Result (regcanc). See Fig. 19.2-4. The message does not include parameters.

AuthenticationRequest Invoke (AUTHRQST). See Fig. 19.2-5. The MS is identified by Par.14 and Par.15 (MIN and MSN), and the invoker includes its MSCID (Par.16) and PC_SSN address (Par.18). Also included are Par.20 RandomVariable (RAND), which is an input to CAVE, and Par.2 AuthenticationResponse (AUTHR), the result of the CAVE execution by the mobile. Authentication can take place when a MS registers, originates, or responds to a page message. On registrations or page responses, the CAVE inputs are MIN, MSN, RAND (received in the invoke), and SSD_A (stored at the AUC). On originations, the invoke also includes Par.11 (Digits), which holds the number dialed by the MS and is used—in lieu of MIN—as an input to CAVE.

AuthenticationRequest Return-Result (authrqst). See Fig. 19.2-5. If AUC determines that the MS is authentic, the message does not include parameters, except when the MS has voice privacy. In that case, the message includes Par.27 VoicePrivacyMask (VPMASK, see Section 12.6.8).

If AUC determines that the MS is not authentic, the message includes Par.15 (DenyAccess).

Return-Error Messages. The error codes in return-error messages, which are common to all IS-MAP operations described in this chapter, indicate why a requested remote operation could not be executed. Some common reasons are

Operation	**Timer**: 6 seconds	
AuthenticationRequest (AUTHRQST)		
Invoke Parameters	**M/O**	**Reference**
AuthenticationResponse (AUTHR)	M	Par.2
Digits (dialed by MS)	O	Par.11
MobileIdentificationNumber (MIN)	M	Par.14
MobileSerialNumber (MSN)	M	Far.15
MSCID (serving MSC or VLR)	M	Par.16
PC_SSN (serving MSC or VLR)	O	Par.18
RandomVariable (RAND)	M	Par.20
Return-Result Parameters	**M/O**	**Reference**
DenyAccess	O	Par.9
VoicePrivacyMask (VPMASK)	O	Par.27
Return-Error Information	**M/O**	**Reference**
Error Code	M	Section 19.5.3
FaultyParameter	O	Par.12

Figure 19.2-5. Parameters in authentication request messages. (From TSB-51. Reproduced with permission of TIA.)

unrecognized MIN (the responder has no record for the received MIN) and unrecognized MSN (the responder's record shows a different MSN for the received MIN). Error codes are listed in Section 19.5.3.

If the error code indicates ParameterError, UnrecognizedParameterValue, or MissingParameter, the message also includes a Par.12 (FaultyParameter) that holds the name (tag) of the parameter that has caused the problem.

19.2.6 Timers

When beginning a transaction (sending an invoke), the invoker starts a timer. If the timer expires and no response has been received, the invoker concludes that a problem has occurred. It then tries to recover, for example, by repeating the invoke. Figures 19.2-2 through 19.2-5 show the timer values for the transactions described above.

19.2.7 Called Addresses of Invokes

We now outline how the initiators of the invokes determine the called party addresses [2].

We first consider the invokes of Fig. 19.2-1.

REGNOT Case (a). The REGNOT concerns a MS whose home CMN is also CMN of MSC-A. MSC-A knows the PC of the HLR of its CMN.

REGNOT Cases (b, c, d). The REGNOT sent by MSC-B concerns a roaming MS. MSC-B knows the PC address of its associated VLR-P.

REGNOT Cases (b, d). VLR-P has to access the HLR of the roaming MS. It uses a global title (GT), in which the address GTA is the MIN of MS, and the translation type is set to TT = 3, to indicate a MIN to HLR translation.

REGCANC Case (d). HLR has the MSCID, and possibly the PC, of VLR-Q, which is currently serving MS. It either uses PC or derives PC from MSCID. VLR-Q has the MSCID, and possibly the PC, of MSC-C, which is currently serving MS.

In conclusion, we consider the invokes of Fig. 19.2-2.

AUTHRQST. MSC-B is serving a roaming MS and sends its AUTHRQST to its associated VLR-P. This register addresses HLR with a global title, as in REGNOT cases (b, d). HLR knows the PC of the AUC in its network. Also, AUC is often implemented as a part of HLR. In this case, HLR passes its AUTHRQST to AUC in a primitive.

19.3 CALLS TO MOBILE STATIONS

Calls to a MS are routed by the PSTN to the "home" MSC of the mobile (Section 19.1.3). This section discusses the operations needed to extend the call setup from the home MSC to the MSC that currently serves the mobile [3].

19.3.1 Temporary Local Directory Numbers

We first introduce the concept of *temporary local directory numbers* (TLDNs). Consider a MSC identified by a particular value of AC-EC. A MS for which MSC is the home MSC has a MIN in the format AC-EC-*XXXX*. When the MS is being called, its MIN is the called party address, and PSTN routes the call to the exchange identified by AC-EC, which is the home MSC of MS.

Each MSC identified by a particular value of AC-EC also has a pool of TLDNs, which also have the AC-EC-*XXXX* format. It temporarily assigns a TLDN to a roaming MS that is being called while being served by MSC.

To distinguish TLDNs from MINs, *XXXX* is divided into two ranges, say, 0000–0999 for TLDNs and 1000–9999 for MINs.

19.3.2 Setup Example

The function of TLDN is illustrated by describing the setup of a call to the MS with MIN = 708-357-8765. The MS homes on MSC-A and is being served by MSC-B, which is identified by AC-EC = 615-443.

The PSTN routes calls to the MS to MSC-A, its "home" switching center. In Fig. 19.3-1, the call setup has reached MSC-A on trunk T_1. MSC-A determines that the received called address is the MIN of a mobile for which it is the home switching center. It then initiates a transaction with the HLR in its network to determine the MSC that is currently serving MS. One result of the operation is that serving MSC-B is informed about an incoming call to the mobile with MIN = 708-357-8765. MSC-B then allocates a TLDN, say, TLDN = 615-443-0089, and records the association of the TLDN with the MIN of the called mobile. Another result of the operation is that MSC-A receives the allocated TLDN.

MSC-A has to route the call to MSC-B. If both MSCs belong to the same network, there may be a CMN trunk group between them, and MSC-A then seizes a trunk in this group.

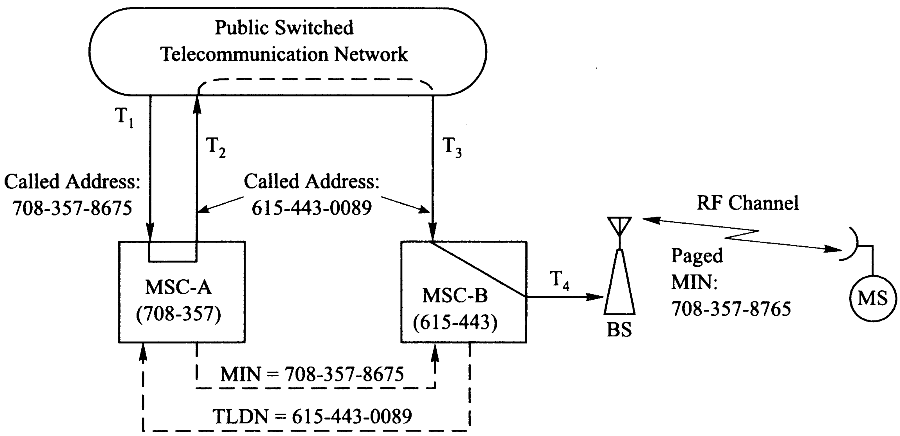


Figure 19.3-1. Connection for a call to mobile (MS).

Figure 19.3-1 assumes that there is no CMN trunk group between the MSCs. MSC-A therefore seizes trunk T_2 to a PSTN exchange, and a connection from MSC-A to MSC-B is set up in the PSTN.

When the setup reaches MSC-B, it determines that the called number 615-443-0089 is one of its TLDNs. It then finds the associated MIN = 708-357-8765 and sends paging messages for that MIN. On receipt of a page response by base station BS, MSC-B seizes trunk T_4 and connects it to T_3. The connection to MS is completed by T_4 and its associated RF voice channel.

MSC-B then discards the record that associates TLDN and MIN and returns TLDN to its pool of temporary local directory numbers.

19.3.3 Operations

The operations involved in setting up a call to a mobile MS are shown in Fig. 19.3-2. Assuming again that MSC-A is the "home" of MS, the PSTN setup reaches MSC-A, which starts a location-request operation with its HLR, sending a LOCREQ invoke. The subsequent operations depend on where the mobile has registered most recently, which is known by HLR.

Case (a). HLR determines that the mobile is currently served by MSC-A and reports this in its locrec return-result. MSC-A then pages the mobile.

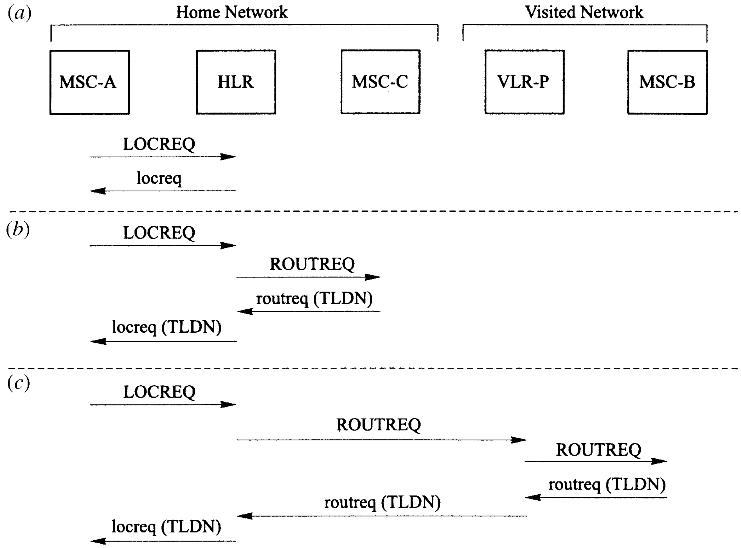


Figure 19.3-2. Operations for terminating calls. (From IS-41.3-B. Reproduced with permission of TIA.)

Case (b). HLR determines that mobile is being served by MSC-C, a switching center in its home network. HLR then initiates a routing-request operation with MSC-C, by sending a ROUTREQ invoke that includes the MIN of the called mobile. MSC-C allocates a TLDN, associates it with the received MIN, and includes TLDN in its return-result (routreq).

The TLDN is also included in the return-result (locreq) sent by HLR. MSC-A then uses the TLDN to route the call to MSC-C. When the setup arrives at MSC-C, it determines the MIN it has associated with the received TLDN and pages the mobile.

Case (c). MS has registered at MSC-B, located in a visited network. In this case, HLR knows the identity of VLR-P, because it has received this information in the REGNOT transaction that was executed when MS registered.

HLR now opens a routing-request transaction with VLR-P, by sending a ROUTREQ invoke, and the VLR opens a routing-request transaction with MSC-B, with a ROUTREQ invoke. The TLDN allocated by MSC-B is included in the routreq return-result sent to VLR-P and HLR, and in the locreq return-result sent to MSC-A. This MSC then routes the call to MSC-B, and the MS is paged.

Case (d). HLR has received a LOCREQ from MSC-A and determines that MS is not currently registered at any MSC. The call thus cannot be set up. HLR indicates this in its locrec return-result, and MSC-A then connects the incoming trunk to a recorded announcement, to inform the calling party.

19.3.4 Message Contents

The main parameters in the invokes and return-results of the operations are listed in Figs. 19.3-3 and 19.3-4. When reading this section, it is helpful to look up the parameter descriptions (19.5.4).

LocationRequest Invoke (LOCREQ). See Fig. 19.3-3. The invoke is sent by the home MSC of the called mobile. Par.11 (Digits) holds the digits received by the MSC, which identify the mobile. The home MSC identifies itself by Par.16 (MSCID) and may include its PC_SSN address (Par.18). Also included is BillingID (Par.5). This is a reference number allocated to the call by the home MSC, for billing purposes. BillingID is also included in the ROUTEREQ invoke to the serving MSC and appears on the call-billing records that are produced by the home and serving MSCs at the end of the call. It is used by the accounting center to calculate the total charge for the call.

LocationRequest Return-Result (locreq). See Fig. 19.3-3. The mobile is identified by Par.14 (MobileIdentificationNumber MIN) and Par.15 (MobileSerial-Number MSN). Par.16 and Par.18 hold the MSCID and PC_SSN address of the serving MSC. Par.11 (Digits) holds the allocated TLDN. If the mobile cannot be accessed, the message includes Par.1 (AccessDeniedReason) instead of Par.11.

Operation	**Timer**: 12 seconds	
LocationRequest (LOCREQ)		
Invoke Parameters	**M/O**	**Reference**
BillingID	M	Par.5
Digits received (MIN)	M	Par.11
MSCID (home MSC)	M	Par.16
PC_SSN (home MSC)	O	Par.18
Return-Result Parameters	**M/O**	**Reference**
AccessDeniedReason	O	Par.1
Digits (TLDN)	O	Par.11
MobileIdentificationNumber (MIN)	M	Par.14
MobileSerialNumber (MSN)	M	Par.15
MSCID (serving MSC or VLR)	M	Par.16
PC_SSN (serving MSC or VLR)	O	Par.18
Return-Error Information	**M/O**	**Reference**
Error Code	M	Section 19.5.3
FaultyParameter	O	Par.12

Figure 19.3-3. Parameters in location request messages. (From IS-41.5-B. Reproduced with permission of TIA.)

RoutingRequest Invoke (ROUTREQ). See Fig. 19.3-4. Par.14 and Par.15 hold MIN and MSN of the called MS. The BillingID, assigned by the home MSC of the called MS, is in Par.5. The identity and PC_SSN address of the home MSC are in Par.16 and Par.18.

Operation	**Timer**: 6 seconds	
RoutingRequest (ROUTREQ)		
Invoke Parameters	**M/O**	**Reference**
BillingID	M	Par.5
MobileIdentificationNumber (MIN)	M	Par.14
MobileSerialNumber (MSN)	M	Par.15
MSCID (home MSC)	M	Par.16
PC_SSN (home MSC)	O	Par.18
Return-Result Parameters	**M/O**	**Reference**
AccessDeniedReason	O	Par.1
Digits (TLDN)	O	Par.11
MSCID (serving MSC)	M	Par.16
PC_SSN (serving MSC)	O	Par.18
Return-Error Information	**M/O**	**Reference**
Error Code	M	Section 19.5.3
FaultyParameter	O	Par.12

Figure 19.3-4. Parameters in routing request messages. (From IS-41.5-B. Reproduced with permission of TIA.)

RoutingRequest Return-Result (routreq). See Fig. 19.3-4. Par.4 and Par.5 hold the MSCID and PC_SSN address of the serving MSC. If the MS can receive calls, the serving MSC has assigned a TLDN, which is in Par.11 (Digits). If the called MS cannot receive the incoming call, for example, when it is already involved in a call, the message includes Par.1 (AccessDenied Reason) in lieu of Par.11.

19.3.5 Called Addresses of Invokes

In conclusion, we outline how the initiators of the invokes of Fig. 19.3-2 determine the called party addresses [2].

LOCREQ Cases (a, b, c). The call for MS has been delivered to its home MSC-A, which knows the point code of the HLR in its network.

ROUTREQ Case (b). Since currently serving MSC-C of MS is in the same network of HLR, this register stores the MSCID, and possibly the point code, of MSC-C. It either derives the PC for the called address from MSCID or uses the stored PC.

ROUTREQ Case (c). Since currently serving MSC-B of MS is not in the network of HLR, this register stores the MSCID, and possibly the point code, of VLR-P. It either derives the PC for the called address from MSCID or uses the stored PC.

VLR-P stores MSCID, and possibly the point code, of currently serving MSC-B and either derives the PC for the called address from MSCID or uses the stored PC.

19.4 OPERATIONS FOR INTERSYSTEM HANDOFF

19.4.1 Introduction

Section 12.1.7 has described the handoff of a mobile that is involved in a call and moves from a cell X to an adjacent cell Y, in the case that both cells are controlled by the same MSC. This section considers a MS handoff when cells X and Y are controlled by different MSCs. These handoffs are known as inter-system handoffs, even though the MSCs usually belong to the same cellular network [5].

In Fig. 19.4-1(*a*), adjacent cells X and Y are controlled by MSC-A and MSC-B, respectively. It is assumed that the MS is in a conversation, using voice channel (VC-U) of base station (BS-X). The connection occupies trunk T_1 (to an exchange in the PSTN network), trunk T_2, and its associated voice channel (VC-U).

MSC-A periodically requests BS-X to monitor the strength of the signal received on VC-U. When the signal strength drops below a certain level, the MSC orders MS to increase its transmit power (Sections 12.1.6 and 12.4.6). However, if MS is already transmitting at its maximum power, it has to be handed off to an adjacent cell. We consider the case that MSC-A has determined that none of its own cells

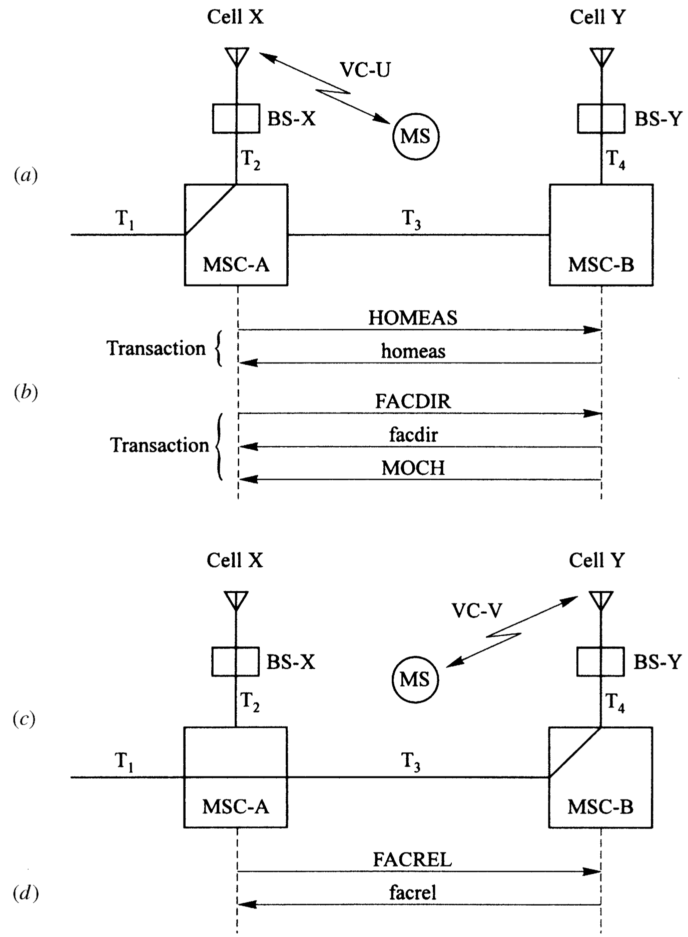


Figure 19.4-1. Handoff of mobile MS from MSC-A to MSC-B: (*a*) configuration prior to handoff, (*b*) handoff, (*c*) configuration after handoff, and (*d*) release of trunk T_2. (From IS-41.2-B. Reproduced with permission of TIA.)

receives a strong enough signal from MS and therefore attempts to hand off the MS to a cell controlled by MSC-B.

Intersystem handoffs require remote operations between the MSCs. Moreover, handoffs from, say, MSC-A to MSC-B require a "handoff" trunk group between the switching centers. The control of the handoff trunk has to be coordinated with other handoff actions and is therefore done as part of a remote operation (which is an exception to the general rule that transactions are not trunk-related). The handoff trunk groups and the inter-MSC trunk groups discussed in Section 19.3 are separate groups, because the latter groups use signaling systems of the types used in fixed networks (MF, ISUP, etc.).

19.4.2 Handoff Example

Figure 19.4-1(*b*) shows the remote operations for the handoff of mobile station MS from MSC-A to MSC-B [4]. The handoff is done in a sequence of two transactions.

Handoff Measurement. The first transaction consists of one operation. It is started by MSC-A, with a handoff measurement invoke (HOMEAS). This requests MSC-B to measure the signal strengths of the MS (on the reverse voice channel of VC-U) received in the cells of MSC-B's service area that are adjacent to cell X. The return-result (homeas) contains a list of "target" cells (cells in which the signal from MS is received with adequate strength and are therefore candidates for the handoff).

Facilities Directive. The second transaction consists of two operations. From the list of target cells, MSC-A selects cell Y. It then seizes a trunk T_3 in its handoff trunk group to MSC-B and starts a facilities-directive operation with a FACDIR invoke that requests MSC-B to set up the connection between T_3 and a voice channel in cell Y.

MSC-B selects an available voice channel VC-V of base station BS-Y. It connects T_3 to trunk T_4 (associated with VC-V) and then sends a facdir return-result. The return-result is sent in a TCAP conversation message, which ends the first operation but continues the transaction (Section 16.1.4).

On receipt of the facdir, MSC-A connects trunk T_1 to trunk T_3, signals the MS to tune to VC-V, and releases T_2 and its associated RF voice channel (VC-U)—see Fig. 19.4-1(*c*).

The facdir return result is sent in a TCAP response package (Section 16.2.1) and the transaction continues.

Mobile on Channel. When MSC-B detects the presence of a signal from MS on VC-V, it starts a mobile-on-channel operation with a MOCH invoke. This operation merely indicates that the handoff is successful and does not require a response. The invoke is therefore sent in a TCAP response package that ends the transaction.

Facilities Release. At the end of the call, trunks T_1, T_3, and T_4 and the RF equipment of associated voice channel VC-V have to be released. In Fig. 19.4-1(*d*), MSC-A has received a disconnect signal on trunk T_1. It then releases trunks T_1 and T_3 at its end and initiates a facilities-release operation, sending a FACREL invoke. This requests MSC-B to release trunk T_3 and the equipment in its part of the connection. MSC-B releases trunks T_3 and T_4 and the RF equipment of VC-V. It then confirms the release with a return-result (facrel), in a TCAP response package that ends the transaction.

If the MS user disconnects first, the MS starts sending signaling tone on voice channel (VC-Y) (Chapter 12). This is detected by MSC-B, which then initiates the release of the facilities.

19.4.3 Message Contents

This section introduces the main parameters in the invokes and return-result for the operations described earlier [5]. The reference numbers (Par.1, etc.) point to parameter descriptions in Section 19.5.4. It is helpful to look up the descriptions when reading this section.

HandoffMeasurementRequest Invoke (HOMEAS). See Fig. 19.4-2. Par.22 (ServingCellID) identifies the cell that currently serves MS. Par.24 (StationClassMark) lists the transmission characteristics of the MS (Section 12.4.8).

If the mobile is currently using an analog voice channel, the invoke includes Par.8 (ChannelData), which contains the channel number, and the attenuation and color codes with which the MS is transmitting.

If the mobile is currently using a digital traffic channel, the invoke also includes Par.10 (DigitalChannelData) and Par.7 (CallMode).

On receiving a HOMEAS invoke, a MSC determines which cells are neighbors to the current serving cell of the MS (from Par.22). These cells can be "target" cells for the handoff. The MSC orders the target cells to measure the signal strengths received from MS.

HandoffMeasurementRequest Return-Result (homeas). See Fig. 19.4-2. After receiving the measurement results, the MSC calculates the "quality" of the received signals. Cells in which the signal quality is above a predetermined level are target cells for the handoff. The return-result is a list of target cells. Each entry consists of a Par.23 (SignalQuality) and a Par.25 (TargetCellID).

Operation HandoffMeasurementRequest	Timer: 7 seconds	
Invoke Parameters	**M/O**	**Reference**
CallMode	O	Par.7
ChannelData (serving)	M	Par.8
DigitalChannelData (serving)	O	Par.10
ServingCellID	M	Par.22
StationClassMark	M	Par.24
Return-Result Parameters	**M/O**	**Reference**
List of target cells. For each cell:		
SignalQuality	M	Par.23
TargetCellID	M	Par.25
Return-Error Information	**M/O**	**Reference**
Error Code	M	Section 19.5.3
FaultyParameter	O	Par.12

Figure 19.4-2. Parameters in handoff measurement messages. (From IS-41.5-B. Reproduced with permission of TIA.)

Operation	Timer: 12 seconds	
FacilitiesDirective		
Invoke Parameters	**M/O**	**Reference**
BillingID	M	Par.5
Callmode	O	Par.7
ChannelData (serving)	M	Par.8
DigitalChannelData (serving)	O	Par.10
InterMSCCircuitID	M	Par.13
MobileIdentificationNumber (MIN)	M	Par.14
MobileSerialNumber (MSN)	M	Par.15
ServingCellID	M	Par.22
StationClassMark	M	Par.24
TargetCellID	M	Par.25
Return-Result Parameters	**M/O**	**Reference**
ChannelData (target)	M	Par.8
DigitalChannelData (target)	O	Par.10
Return-Error Information	**M/O**	**Reference**
Error Code	M	Section 19.5.3
FaultyParamcter	O	Par.12

Figure 19.4-3. Parameters in facilities directive messages. (From IS-41.5-B. Reproduced with permission of TIA.)

FacilitiesDirective Invoke (FACDIR). See Fig. 19.4-3. Par.14 and Par.15 (MIN and MSN) identify the mobile that will be handed off. Par.8, Par.22, and Par.24 (and possibly Par.7 and Par.10) of the HOMEAS invoke are included again. The target cell (selected from the list in the homeas return-result) and the seized inter-MSC trunk are in Par.25 and Par.13. The BillingID (Par.5) is included for billing purposes.

FacilitiesDirective Return-Result (facdir). See Fig. 19.4-3. A MSC receiving a FACDIR invoke allocates an analog or digital channel in the selected target cell and responds with a return-result that includes Par.8. If the allocated channel is a digital traffic channel, Par.10 is also included. These parameters contain information required by the mobile and appear in the handoff message to the mobile that is sent by old serving MSC (Section 12.4.6).

MobileOnChannel Invoke (MOCH). See Fig. 19.4-4. This indicates that the new serving MSC has verified the presence of the signal from MS on the new channel. No parameters are included.

FacilitiesRelease Invoke (FACREL). See Fig. 19.4-5. This indicates that the invoking MSC has released the inter-MSC trunk at its end. Par.14 (MIN) and Par.13 (InterMSCCircuitID) identify the mobile and the inter-MSC trunk. Par.21 (ReleaseReason) indicates why the connection is being released.

Operation	Timer: none	
MobileOnChannel		
Invoke Parameters	**M/O**	**Reference**
None		

Figure 19.4-4. Mobile on channel invoke. *Note*: No return-result or return-error is sent in this operation. (From IS-41.5-B. Reproduced with permission of TIA.)

Operation	Timer: 4-15 seconds	
FacilitiesRelease		
Invoke Parameters	**M/O**	**Reference**
InterMSCCircuitID	M	Par.13
MobileIdentificationNumber	M	Par.14
ReleaseReason	M	Par.21
Return-Result Parameters	**M/O**	**Reference**
None		
Return-Error Information	**M/O**	**Reference**
Error Code	M	Section 19.5.3
FaultyParameter	O	Par.12

Figure 19.4-5. Parameters in facilities release messages. (From IS-41.5-B. Reproduced with permission of TIA.)

FacilitiesRelease Return-Result (facrel). See Fig. 19.4-5. This acknowledges the receipt of FACREL and indicates that the responding MSC has cleared the inter-MSC trunk at its end. No parameters are included.

19.4.4 Called Addresses of Invokes

In conclusion, we outline how the initiators of the invokes of Fig. 19.4-1 determine the called party addresses [2].

HOMEAS and FACDIR. Serving MSC-A has stored information on the point codes of MSCs that have cells that are neighbors of cell X, which currently communicates with the mobile. The example assumes that MSC-B is one of these MSCs. MSC-A thus uses the PC of MSC-B as the called address for its HOMEAS and FACDIR invokes.

MOCH. MSC-B has received the PC of MSC-A in the HOMEAS and FACDIR invokes. It thus knows the PC of MSC-A and uses it in its MOCH invoke.

FACREL. After the transactions of Fig. 19.4-1(*b*), the MSCs know each other's PCs. In the example, MSC-A initiates the release of the connection and uses the point code of MSC-B as called address for its invoke.

19.5 IS-MAP FORMATS AND CODES

The coding principles of Section 16.3 apply, except that IS-MAP documents use the term *parameter* in lieu of *data element*. This section lists the formats and codes of the parameters in the messages discussed in Sections 19.2–19.4 [2].

19.5.1 Tags (Identifiers)

Operation code and error code tags are coded as “private TCAP”:

								← **Bits**
	H	**G**	**F**	**E**	**D**	**C**	**B**	**A**
Operation code identifier	1	1	0	1	0	0	0	1
Error code identifier	1	1	0	1	0	1	0	0

19.5.2 Operation-Code Contents

The contents fields of operation codes consist of two octets. Octet 1 is coded as

← **Bits**

H	**G**	**F**	**E**	**D**	**C**	**B**	**A**
0	0	0	0	1	0	0	1

This indicates that the operation belongs to the mobile applications part IS-MAP operations family.

Octet 2 specifies a particular operation—see Table 19.5-1.

TABLE 19.5-1 Operation-Code Specifiers

Decimal Value	Operation
1	HandoffRequestMeasurement
2	FacilitiesDirective
3	MobileOnChannel
5	FacilitiesRelease
13	RegistrationNotification
14	RegistrationCancellation
15	LocationRequest
16	RoutingRequest
28	AuthenticationRequest

Source: IS-41.5-B. Reproduced with permission of TIA.

19.5.3 Error-Code Contents

The one-octet contents field indicates why a MSC, HLR, or VLR cannot execute a requested operation. Bits H,G,F,E are set to 1,0,0,0. Bits D to A contain an integer that indicates a specific error, for example:

Value	Meaning
1	*UnrecognizedMIN*. The received mobile identification number is not currently served by the HLR, VLR, or MSC.
2	*UnrecognizedMSN*. A HLR or VLR has received a combination of a MIN and a MSN and, according to its records, the received MSN is not associated with the received MSN.
3	*MIN/HLRMismatch*. The received mobile identification number is not known at the HLR.
4	*OperationSequenceProblem*. The requested operation is not allowed in the current state of the call.
8	*ParameterError*. Parameter is not expected or is coded incorrectly.
10	*UnrecognizedParameterValue*. Parameter value is coded properly but has an unrecognized value.
12	*MissingParameter*. Not all expected message parameters have been received.

When the error code is 8, 10, or 12, the return-error component also includes one or more Par.12 (FaultyParameter), which specify the parameter(s) that caused the problem.

19.5.4 Parameter Descriptions and Formats

The parameter tags (identifiers) are coded as "context-specific" (bits H,G = 1,0; see Section 16.3.2) and are listed in Table 19.5-2. The contents fields of the parameters are described below.

Par.1 AccessDeniedReason (one octet). This is an integer indicating why the mobile should not receive (terminate) a call:

Value	Meaning
1	Mobile identification number not recognized
2	Mobile inactive
3	Mobile busy
4	Terminations to mobile are denied

Par.2 AuthenticationResponse (AUTHR; three octets). This contains the 18-bit result of the CAVE algorithm executed by the MS (Section 12.6.7).

TABLE 19.5-2 Parameter Type Identifiers (Tags)

Reference	Parameter Name	← Bits HGFE	DCBA
Par.1	AccessDeniedReason	1001	0100
Par.2	AuthenticationResponse (AUTHR)	1001	1111
		0010	0011
Par.3	AuthorizationDenied	1000	1101
Par.4	AuthorizationPeriod	1000	1110
Par.5	BillingID	1000	0001
Par.6	CallingFeaturesIndicator	1001	1001
Par.7	CallMode	1001	1101
Par.8	ChannelData	1000	0101
Par.9	DenyAccess	1001	1111
		0011	0010
Par.10	DigitalChannelData	1001	1100
Par.11	Digits	1000	0100
Par.12	FaultyParameter	1001	1010
Par.13	InterMSCCircuitID	1000	0110
Par.14	MobileIdentificationNumber (MIN)	1000	1000
Par.15	MobileSerialNumber (MSN)	1000	1001
Par.16	MSCID	1001	0101
Par.17	OriginationIndicator	1001	0111
Par.18	PC_SSN	1001	1111
		0000	0010
Par.19	QualificationInformationCode	1001	0001
Par.20	RandomVariable (RAND)	1001	1111
		0010	1000
Par.21	ReleaseReason	1000	1010
Par.22	ServingCallID	1000	0010
Par.23	SignalQuality	1000	1011
Par.24	StationClassMark	1000	1100
Par.25	TargetCellID	1000	0011
Par.26	TerminationRestrictionCode	1001	1000
Par.27	VoicePrivacyMask (VPMASK)	1001	1111
		0011	0000

Note: Tags of Par.2, Par.9, Par.18, Par.20, and Par.27 occupy two octets—see Section 16.2.3.

Source: IS-41.5-B. Reproduced with permission of TIA.

Par.3 AuthorizationDenied (one octet). This integer indicates why the mobile should not receive service:

Value	Meaning
1	Delinquent account
2	Invalid mobile serial number
3	Stolen unit
5	Mobile identification number not recognized
6	Unspecified

Par.4 AuthorizationPeriod (two octets). The parameter indicates that the MSC is authorized to serve the mobile station for a certain period. The length of the period is indicated by two integers. Octet 1 specifies a unit of time:

Value	Meaning
0	Per call authorization
1	Hours
3	Days
5	Weeks

Octet 2 indicates the number of time units (1–255) of the authorization period. When the value of octet 1 is 0, the value of octet 2 is also 0.

Par.5 BillingID (seven octets). This is a reference number for billing purposes. It is assigned to a call by the "anchor" MSC (the first MSC involved in the call).When a call involves several MSCs, because of roaming or intersystem handoffs, the BillingID is passed on to the next serving MSC. All MSCs involved in the call include the BillingID in their billing records. In this way, the billing records for the call can be correlated.

In Fig. 19.5-1, octets 1–3 identify the anchor MSC: octets 1 and 2 hold the 15-bit system identification (SID) that identifies a cellular system (Chapter 12), and octet 3 holds the switch number (SWNO) that identifies a MSC within the cellular system.

Octets 4–6 contain the billing number.

Par.6 CallingFeaturesIndicator (two octets). This contains a number of two-bit indicators, associated with calling features that can be offered to mobiles:

Indicator	Feature
a_1, a_2	Call forwarding—unconditional
b_1, b_2	Call forwarding if MS is busy
c_1, c_2	Call forwarding if MS is busy
d_1, d_2	Call waiting
e_1, e_2	Three-way calling
f_1, f_2	Call delivery
g_1, g_2	Voice privacy

The indicator values define the status of each feature for a particular mobile:

Value	Feature
0,1	Not authorized for MS
1,0	Authorized but deactivated by MS
1,1	Authorized and activated by MS

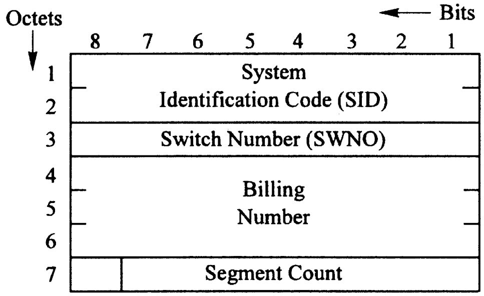


Figure 19.5-1. Format of Par.5 (BillingID). (From IS-41.5-B. Reproduced with permission of TIA.)

Par.7 CallMode (one octet). This indicates the mode of the current call: analog or digital.

Par.8 ChannelData (zero or three octets). If a mobile is using an analog voice channel, the contents field has three octets and contains the color code (SCC) of the channel, the most recent mobile attenuation code (VMAC) sent to the mobile, and the channel number (CHAN) of the channel (see Section 12.4.8 for definitions of SCC, VMAC, and CHAN). If a mobile is using a digital traffic channel, the parameter appears in the message, but the length of its contents field is 0 octet.

Par.9 DenyAccess (one octet). This is an integer indicating the reason why an authentication center has determined that the mobile is not authentic and should not be allowed to access the mobile network. The octet is usually coded "1" (reason not specified).

Par.10 DigitalChannelData (five octets). If a mobile is using an IS-54 TDMA traffic channel in a digital channel (Section 12.6.2), this parameter contains:

Channel Number (*CHAN*). This identifies a 48.6-kb/s digital channel (Section 12.6.2).

Time-Slot Indicator (*TSI*). A digital channel contains three TDMA traffic channels. TSI identifies the time-slot pair (1 and 4, 2 and 5, or 3 and 6) used by the traffic channel.

Digital Color Verification Code (*DVCC*). This is the color code of a digital channel—see Section 12.6.2.

Digital Mobile Attenuation Code (*DMAC*). This is the digital counterpart of VMAC (voice mobile attenuation code)—see Section 12.4.8.

Par.11 Digits (variable length). This holds a string of BCD coded digits—see Fig. 19.5-2. The integer in octet 1 indicates the type of the digit string:

Value	Meaning
1	Dialed number
2	Routing number
3	Destination number
4	Carrier identification code

Octet 2 is not used. The code in octet 3 indicates that the number has the format of the U.S. telephone numbering plan, and that the digits are BCD coded. The integer in octet 4 indicates the number of digits. The digits are located in octets 5 through *n*.

Par.12 FaultyParameter (one or two octets). This contains the parameter tag (Table 19.5-2) of a missing, incorrect, or unexpected parameter.

Par.13 InterMSCCircuitID (two octets). This identifies an inter-MSC trunk. Octets 1 and 2 contain, respectively, the group and member number of the trunk.

Par.14 MobileIdentificationNumber (MIN). This consists of five octets. Each octet contains two BCD coded digits of a ten-digit mobile identification number (Section 12.3).

Par.15 MobileSerialNumber (MSN). This consists of four octets, containing the 32-bit serial number of a mobile (Section 12.3).

Par.16 MSCID (three octets). This identifies the MSC or VLR that has opened the transaction. Octets 1 and 2 hold the SID (system ID) that identifies a cellular mobile network (Section 12.3). Octet 3 identifies a MSC or VLR in that network.

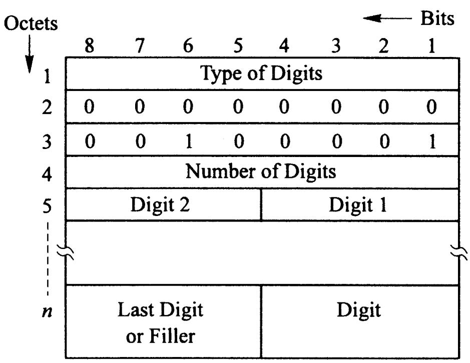


Figure 19.5-2. Format of Par.11 (Digits). (From IS-41.5-B. Reproduced with permission of TIA.)

Par.17 OriginationIndicator (one octet). This integer indicates the destinations allowed to be called by the mobile:

Value	Meaning
2	No originations allowed
3	Local calls only
4	Calls to selected area codes only
5	Calls to selected area codes and local calls
6	Calls to area codes of all destinations in world zone 1 (see Section 1.3.3) and local calls

Par.18 PC_SSN (five octets). This identifies a MAP subsystem in a signaling point. Octet 1 indicates the subsystem type:

Value	Meaning
1	Serving MSC
2	Home MSC
4	Home location register (HLR)
5	Visitor location register (VLR)
6	Authentication center (AUC)

Octets 2–4 hold the point code (PC); octet 5 holds the subsystem number (SSN).

Par.19 QualificationInformationCode (one octet). This integer indicates the requested data about a mobile:

Value	Meaning
1	No information requested
2	Validation only
3	Validation and service profile
4	Service profile only

Par.20 RandomVariable (RAND; four octets). This holds the 32-bit random number used by the MS as input to the CAVE algorithm (Section 12.6.7).

Par.21 ReleaseReason (one octet). This is an integer indicating why the inter-MSC trunk is being released:

Value	Meaning
0	Not specified
1	Call ended, clear-forward received
2	Call ended, clear-back received

Par.22 ServingCellID (two octets). This is an integer that identifies the cell that is currently serving the mobile.

Par.23 SignalQuality (one octet). This is an integer that indicates the quality (strength) of the received signal in a reverse voice or traffic channel:

Value	Meaning
1–8	Signal too weak
9–245	Acceptable signal
245–255	Signal too strong (may cause cochannel interference)

Par.24 StationClassMark (one octet). This parameter contains the transmission characteristics of the mobile (Section 12.4.8).

Par.25 TargetCellID (two octets). This is an integer that identifies a cell to which a mobile can be handed off.

Par.26 TerminationRestrictionCode (one octet). This is an integer indicating the call types which the mobile station is allowed to receive:

Value	Meaning
1	No terminations allowed
2	No restrictions on terminations
3	Restricted to sent-paid calls (calls billed to the calling party)

Par.27 VoicePrivacyMask (VPMASK; 66 octets). This is a 528-bit field that holds two 260-bit masks (one for each direction of transmission). Used to scramble and unscramble the 260-bit user data blocks for IS-54 mobiles that have the voice privacy feature (Section 12.6.8).

19.6 INTRODUCTION TO GSM-MAP

The remainder of this chapter briefly describes GSM-MAP, the mobile application part defined by the Telecommunication Standardization Sector of the International Telecommunication Union (ITU-T) and the European Telecommunications Standards Institute (ETSI).

GSM-MAP supports communications in mobile networks of the Global System for Mobile Communications (GSM—see Section 12.7), which are installed in a large number of nations.

As in IS-MAP, the main functions of GSM-MAP support MS registration, MS roaming, MS authentication, and MS handover (the European term for handoff). International roaming is allowed. A customer who has a GSM service provider in one nation can roam in GSM networks of other nations.

19.6.1 GSM-MAP Terms

A number of GSM-MAP terms and acronyms are identical to those of IS-MAP, for example, mobile station (MS), mobile switching center (MSC), home location register (HLR), visitor location register (VLR), authentication center (AUC), roamer, and location area. Some differences in terminology are outlined below [6].

Public Land Mobile Network (PLMN). This is the equivalent of cellular mobile network (CMN).

Home PLMN. This is the PLMN selected by the MS owner to provide mobile communications. It is the equivalent of the home CMN.

Visited PLMN. This is the PLMN that currently serves a roaming MS. It is the equivalent of visited CMN.

Serving MSC. This is the MSC that is currently serving the MS.

Gateway MSC (GMSC). This is the MSC to which the fixed (PSTN/ISDN) network delivers a call to the MS. It is nearly equivalent to the home MSC. In U.S. mobile networks, calls to a particular MS are always delivered to its home MSC, which is determined by analyzing the MIN of the mobile.

GSM networks do not have home MSCs. Instead, the operator of a PLMN designates a number of MSCs as gateway MSCs (GMSCs), and calls to a particular MS are delivered by the fixed network to one of the GMSCs in the home PLMN of the mobile.

Mobile Station ISDN Number (MSISDN). This is the number dialed by a subscriber when calling the MS. It is used by the fixed network to route calls for MS to a nearby gateway MSC in the home PLMN of MS.

Mobile Station Roaming Number (MSRN). This is the equivalent of the temporary local directory number (TLDN).

International Mobile Equipment Identity (IMEI). This is the equivalent of the mobile serial number (MSN).

19.6.2 Equipment Entities

The equipment entities in a PLMN that are involved in GSM-MAP transactions are shown in Fig. 19.6-1 [6].

Mobile Switching Centers (MSC). In this example, the PLMN has two MSCs, each of which serves the mobiles in its service area.

Home Location Register (HLR). This register stores semipermanent and temporary data about the mobiles for which PLMN is the home network.

Visitor Location Register (VLR). The register has records on the mobiles that are currently being served by its associated MSCs. In this example, VLR-P and VLR-Q are associated with MSC-A and MSC-B, respectively.

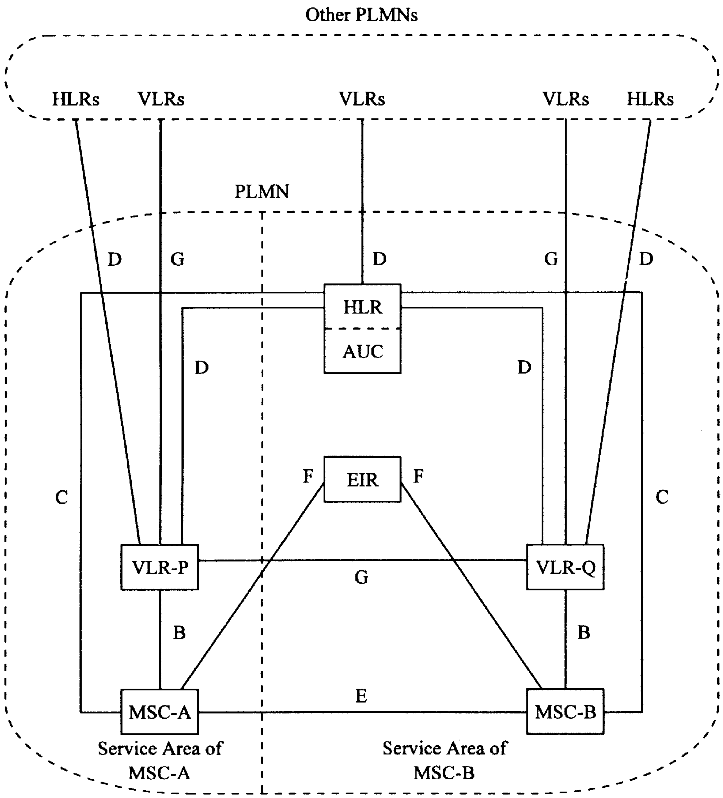


Figure 19.6-1. Equipment entities and interfaces. (From GSM.09.02-Version 4.9.0. Courtesy of ETSI.)

Authentication Center (AUC). This register stores authentication and voice privacy information about the mobiles for which PLMN is the home network.

Equipment Identity Register (EIR). This register stores the *international mobile equipment identity* (IMEI) of all mobile equipment (Section 12.7.4) for which PLMN is the home network.

19.6.3 Interfaces

We next explore the GSM-MAP interfaces between the equipment entities of Fig. 19.6-1 [6]. There are some differences from the IS-MAP interfaces.

The interface between a MSC and its associated VLR is known as the “B” interface. ETSI considers the B interface as “internal” and does not define it. The message transfer does not involve the signaling network.

The C interface is between a gateway MSC and the HLR in its PLMN. The D interface is between a VLR and a HLR, in the same or different PLMNs.

The E interface is between adjacent MSCs in a PLMN and is used in handover operations.

The F interface is an intra-PLMN interface, between a MSC and the EIR.

The G interface is between VLRs, in the same or in different PLMNs. This interface does not exist in IS-MAP.

In GSM-MAP, the AUC is an integral part of HLR, and no interface is defined.

19.6.4 Numbering Plans

In the United States mobile networks use the numbering plan that was already in place for fixed networks. The MIN of a mobile has the format of national numbers (Section 1.2.1).

Nations with GSM mobile networks have a fixed network numbering plan and a separate PLMN numbering plan.

In the fixed (PSTN/ISDN) network, a mobile is identified by a MSISDN, which has the format defined in ITU-T Rec. E.164 [7]—see Section 1.2. MSISDN is an international number, consisting of country code CC and national number NN.

The mobile station roaming number (MSRN), which is used by the fixed network to extend the call setup to the MSC currently serving the called MS, is a national or international E.164 number.

In the PLMN, a MS is idenfied by either an *international mobile station identity* (IMSI) or by the combination of a *location area identity* (LAI) and a *temporary mobile station identity* (TMSI) (Section 12.7.4).

The IMSI format has been defined in ITU-T Rec. E.212 [8]:

$$\text{IMSI} = \text{MCC-MNC-MSIN}$$

where the combination of the mobile country code MCC and the mobile network code MNC identifies the home PLMN of MS, and MSIN is the mobile station

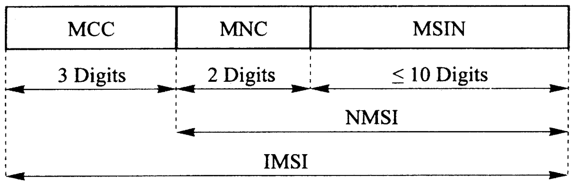


Figure 19.6-2. Format of IMSI. (From GSM.09.02-Version 4.9.0. Courtesy of ETSI.)

identity that identifies a mobile in that PLMN—see Fig. 19.6-2. IMSI is an international number that uniquely identifies a mobile worldwide. In its home country, a MS can be uniquely identified by a *national mobile station identity* (NMSI):

$$\text{NMSI} = \text{MNC-MSIN}$$

MSISDN, IMSI, and NMSI permanently identify a MS. In addition, the combination of a *location area identity* (LAI) and *temporary mobile station identity* (TMSI) identifies a MS during the time it is served by the VLR that covers the location area. The format of LAI is

$$\text{LAI} = \text{MCC-MNC-LAC}$$

where MCC-MNC identifies a particular PLMN, and the *local area code* (LAC) represents a location in the PLMN.

TMSI is a 32-bit binary number that identifies the MS while it is operating in a particular location area.

Because of the multiple MS identities and numbering plans, GSM-MAP requires a number of operations that are not encountered in IS-MAP.

19.6.5 Information in SIM, HLR, VLR, EIR, and AUC

The most important parameters stored in the subscriber identity module (SIM) of a MS (Section 12.7.4), and in the various GSM network entities, are listed below [9]. Semipermanent and temporary parameters are denoted by (S) and (T).

Information in SIM

- IMSI (S)
- TMSI (T)
- LAI (T)
- Ki (S), authentication key (Section 12.8.4)
- Kc (T), cipher key (Section 12.8.4)

A MS updates its temporary parameters on command from the serving MSC.

Information in HLR. The HLR of a PLMN has records for all MSs whose home network is PLMN.

- MSISDN (S) of MS
- IMSI (S) of MS
- Originating and terminating service profile of MS (S)
- Address of the VLR associated with the MSC that is currently serving MS (T)

Information in VLR. This register has a record for each MS currently served by one of its associated MSCs.

- MSISDN (T)
- Originating and terminating service profile of MS (T)
- IMSI (T)
- TMSI in the current location area (T)
- LAC of the current MS location area (T)
- MSRN that is currently assigned to the MS (T)

Information in AUC. The AUC in a PLMN has records with authentication and privacy information for all mobiles whose home network is PLMN.

- MSISDN (S)
- Ki (S)
- Kc (T)

19.6.6 SCCP Addresses

GSM-MAP messages are transferred by the message transfer part (MTP) and the signaling connection control part (SCCP) of the SS7 signaling networks that also transfer messages for fixed-network applications.

GSM-MAP uses the connectionless services of SCCP (Section 15.3.3).

We now explore the SCCP addresses of GSM-MAP entities. The application service entities (ASE) at a signaling point are addressed by subsystem numbers (SSNs):

0000 0110	Home location register
0000 0111	Visitor location register
0000 1000	Mobile switching center
0000 1001	Equipment identity register

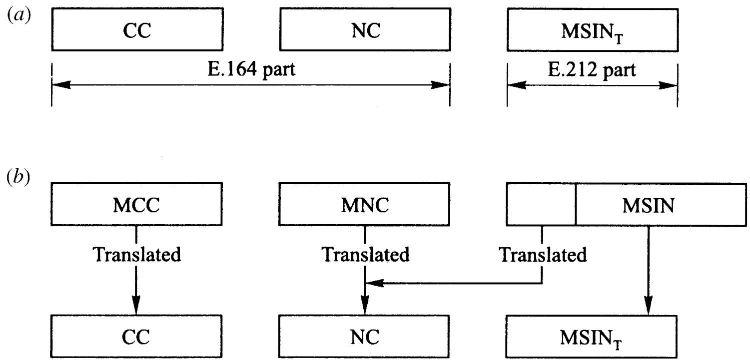


Figure 19.6-3. Format and derivation of E.214 global title. (From GSM.09.02-Version 4.9.0. Courtesy of ETSI.)

In SCCP called (CDA) and calling (CGA) addresses, SSN is always present. In addition, the addresses include a point code (PC) and/or a global title (GT). This depends on locations of the entities involved in the transaction:

- If both entities are in the same PLMN, the address is a point code.
- If both entities are in the same country but in different PLMNs, the address is a global title with the format of an E.164 number, according to the numbering plan for the fixed network in that country.
- If the entities are in different countries, the address is a global title with a format defined in ITU-T Rec. E.214 [10].

The format is shown in Fig. 19.6-3(*a*). It consists of an E.164 part (country code CC and network code NC) and an E.212 part ($MSIN_T$), which represents the leading digits of MSIN.

The derivation of this address from the IMSI of a mobile is shown in Fig. 19.6-3(*b*). Country code CC is derived from mobile country code MCC. Network code NC is derived from mobile network code MNC and the leading digits of MSIN. $MSIN_T$ is truncated if necessary, in order to limit the length of the address to 15 digits.

19.6.7 Derivation of Called Address

The entity that initiates a transaction has to derive the address of the called entity from an input datum. We explore the address derivations in a number of situations that frequently occur in GSM-MAP.

A VLR initiates a transaction with the HLR of a mobile identified by IMSI. VLR derives the called address from IMSI.

A gateway MSC initiates a transaction with the HLR of a mobile identified by MSISDN. In this situation, the GMSC, mobile, and HLR belong to the same PLMN. Most PLMNs have one HLR, and the called address (a point code) is fixed. If a PLMN has several HLRs, GMSC derives the address from MSISDN.

VLR_1 initiates a transaction with VLR_2, which covers a particular location area (LAI). In this case, VLR_1 derives the address of VLR_2 from LAI. A VLR has no access to a VLR in another country.

A VLR initiates a transaction with its associated MSC, or vice versa. The interface between these two entities is internal (B interface), and addressing is not required.

A HLR initiates a transaction with the VLR serving a MS. This situation occurs only if the VLR initiated a previous transaction concerning MS with HLR. During the previous transaction, HLR has stored the VLR address in its record on MS and uses this address when it initiates its transaction.

19.6.8 GSM-MAP Transactions and Operations

As in IS-MAP, the GSM-MAP transactions consist of one operation and require two messages. GSM-MAP uses the terms of ITU-T TCAP: a Begin message (with one invoke) initiates the transaction, and an End message (with a return-result or return-error) ends the transaction.

In the sections that follow, the messages of a transaction are shown as in the example of Fig. 19.6-4, where ASE-A requests ASE-B to provide a roaming number for the mobile identified by IMSI. The Begin message is denoted by MAP_ SEND_ ROAMING_NUMBER, which indicates the invoked operation. The End message is denoted by the same character string, followed by *ack* [9].

In what follows, the most important parameters in invokes and return-results are shown in parentheses, below the respective messages.

19.7 OPERATIONS RELATED TO LOCATION UPDATING

This section discusses location updating, which is the counterpart of registration notification in IS-MAP (see Section 19.2). Its purposes are to verify that the MS qualifies for service (this may include MS authentication, etc.) and to update the HLR regarding the current MS location. Also, if MS registers in a location area covered

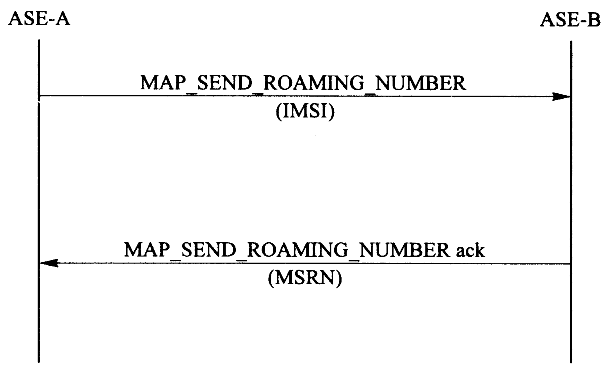


Figure 19.6-4. Messages in a GSM-MAP transaction.

by a “new” VLR, the VLR needs to establish a MS record and therefore has to interrogate the HLR [9,11].

19.7.1 Description of Operations

The MAP operations for location updating and the most important parameters in their invokes and return-results are outlined below. The operations occur in the order they are executed during a typical location update.

In what fallows, the “HLR of MS” is the home location register in the home PLMN of MS that has a record for MS (Section 19.6.5). MSC is the mobile switching center that has received a location update (LU) request from MS, and VLR is the visitor location register associated with that MSC. PVLR is the “previous” visitor location register, which has been serving the MS up to this point.

MAP_UPDATE_LOCATION_AREA. The operation is invoked by a MSC that has received a LU_request from a MS. It requests VLR to establish a record for MS if it does not already have one and, if desired by VLR, to authenticate the MS, and/or to allocate a new TMSI. The acknowledgment by VLR indicates that it has successfully performed these tasks.

MAP_UPDATE_LOCATION. The operation is invoked by a VLR whose MSC has received a LU_request from a mobile MS, and which has no record on MS. It requests the HLR of MS to enter the VLR address as the new current location of MS in its MS record.

MAP_CANCEL_LOCATION. The operation is invoked by the HLR of MS whose MS record includes a VLR address and has received an update_location invoke about MS. The cancel_location invoke is made to PVLR, which is identified by its address. The invoke requests PVLR to erase its MS record. After doing so, PVLR returns an acknowledgment.

MAP_INSERT_SUBSCRIBER_DATA. This operation is invoked by the HLR of mobile MS, after receiving an update_location invoke from a VLR. The invoke includes information items in the MS record at HLR (Section 19.6.5). VLR establishes a MS record that contains the information and then acknowledges the invoke.

MAP_SEND_IDENTIFICATION. The operation is invoked by the VLR associated with a MSC that has received a LU_request from mobile MS in which MS has identified itself by the $TMSI_S$ and LAI_S that are stored in its SIM. The invoke, which includes these parameters, requests PVLR to provide the IMSI of MS. PVLR includes IMSI in its acknowledgment.

MAP_PROVIDE_IMSI. The operation is invoked by the VLR associated with the MSC that has received a LU_request from a MS in which the MS is identified by

$TMSI_S$ and LAI_S. The invoke requests MSC to provide the IMSI of the mobile. MSC requests the IMSI from MS and includes it in its return-result.

MAP_SEND_AUTHENTICATION_INFO. The GSM authentication procedures at the authentication center (AUC) and the mobile are shown in Fig. 19.7-1. The GSM authentication algorithm (A) has two inputs: random number RAND and authentication key Ki.

At AUC, a random number source generates a new RAND for each authentication. AUC stores a Ki_A for all mobiles that are "at home" in its PLMN. The MS to be authenticated is identified by IMSI. AUC finds the Ki_A associated with IMSI and then performs the algorithm. The results are the *signed result* ($SRES_A$) and the cipher key (Kc_A). Kc is used for data encryption on the radio interface.

At MS, the algorithm is performed using the same RAND value and with the authentication key Ki_S (stored in its SIM). The result is $SRES_S$.

MS authentication is the responsibility of a VLR. It requests AUC/HLR and MS to perform the algorithm and then compares the two results.

The invoke requests HLR/AUC to perform the algorithm for a mobile identified by its IMSI. The acknowledgment includes $SRES_A$, Kc, and the RAND value that was used.

MAP_AUTHENTICATE. The operation is invoked after VLR has obtained the authentication results of AUC. The invoke requests MSC to obtain authentication results from MS and includes the RAND value used by AUC. MSC then requests MS to perform the authentication algorithm. After receiving the MS response, MSC acknowledges and includes $SRES_S$.

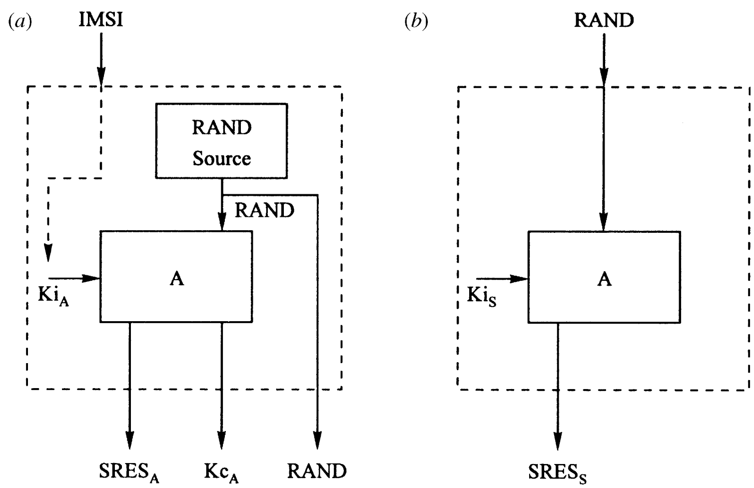


Figure 19.7-1. Authentication algorithm. (From GSM.09.02-Version 4.9.0. Courtesy of ETSI.)

VLR now compares $SRES_A$ and $SRES_S$. If the values match, the MS must have the proper Ki_s and is therefore authentic. Otherwise, VLR informs MSC, which then sends a release message to MS.

MAP_SET_CIPHERING_MODE. This operation is invoked by VLR. It requests MSC to inform MS, and the BS that is serving MS, to initiate the ciphering (encryption) of information on the radio (RF) channel (Um interface). The invoke includes the cipher key (Kc), which is used by BS and MS to construct the encryption masks.

MAP_FORWARD_NEW_TMSI. The operation is invoked by a VLR that has decided to allocate a new TMSI to a MS. The invoke includes the new TMSI and requests the associated MSC to forward the new TMSI to MS. MSC then informs MS and, after receiving confirmation, acknowledges the invoke.

19.7.2 Location Updating Examples

We now explore a number of examples [9]. In the figures of this section, MSC receives a location updating request from a MS, and VLR is associated with this MSC. HLR is the home location register of MS. PVLR is the "previous" visitor location register, associated with the MSC that had been serving MS up to this point.

In Fig. 19.7-2, MS does a location update in a location area covered by the same VLR that was involved in the previous location update of MS. VLR thus already has a record on MS. The update_location_area invoke includes the LAI_S and $TMSI_S$, stored in the SIM of MS, and LAI_B, the location area identity of the base station (BS) that received the LU_request. If LAI_S and LAI_B do not match, VLR updates its MS record, entering LAI_B as the new MS location area. It then acknowledges, and MSC confirms the MS request with a LU_confirm message, which includes LAI_B. MS then updates its SIM, replacing LAI_S by LAI_B.

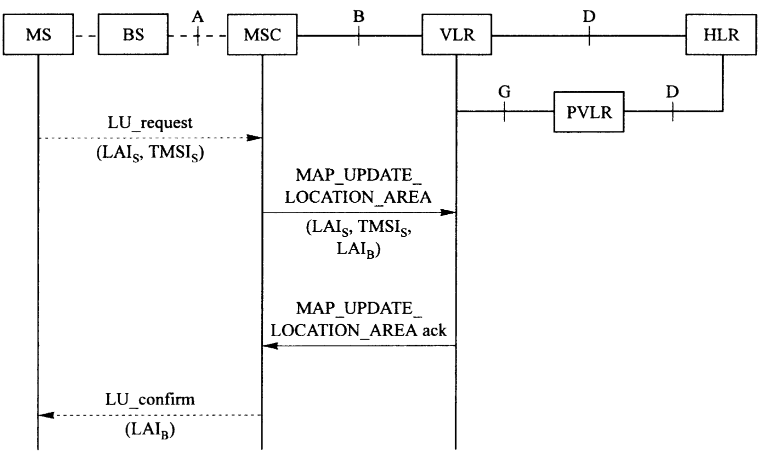


Figure 19.7-2. Operations for location area updating. Mobile is known at VLR. (From GSM.09.02-Version 4.9.0. Courtesy of ETSI.)

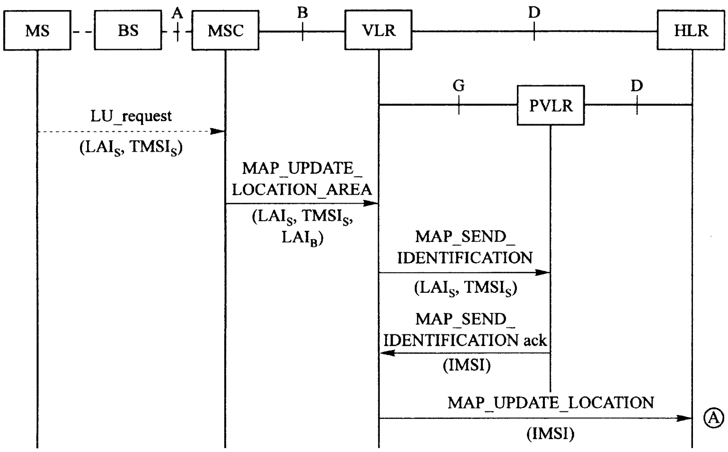


Figure 19.7-3. Initial operations for location area updating. Mobile is not known at VLR and is identified by LAI$_S$ and TMSI$_S$. (From GSM.09.02-Version 4.9.0. Courtesy of ETSI.)

The next examples consider MS location updates in a location area covered by a new VLR. The VLR now has to establish a MS record. We first explore the operations that enable VLR to interrogate the HLR of MS.

In Fig. 19.7-3, MS sends a LU_request that again includes LAI$_S$ and TMSI$_S$. VLR finds no record for MS and has to interrogate HLR to obtain information for its new record. However, the HLR of MS cannot be determined from LAI$_S$ and TMSI$_S$.

Therefore, VLR initiates a send_information operation with PVLR, requesting the IMSI of the MS identified by LAI$_S$ and TMSI$_S$. VLR derives the address of PVLR from LAI$_S$. PVLR finds the associated MS record and includes the IMSI in its acknowledgment.

VLR then derives the address of HLR from IMSI and initiates an update_location operation (A) with the HLR.

In Fig. 19.7-4, MS has identified itself by the IMSI$_S$ stored in its SIM. This is done when the SIM is used for the first time, and when SIM has lost the LAI$_S$ and/or TMSI$_S$. In this case, VLR can access HLR immediately (A).

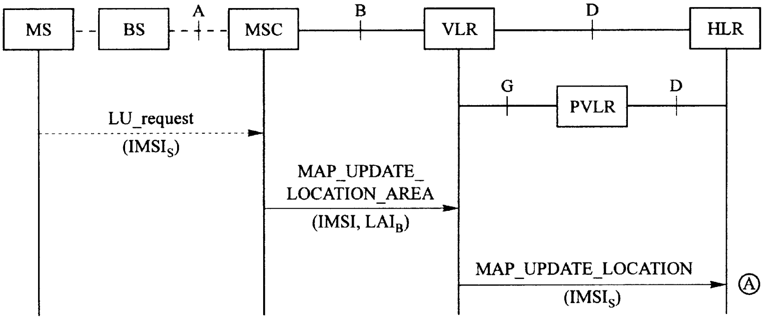


Figure 19.7-4. Initial operations for location area updating. Mobile is not known at VLR and is identified by IMSI. (From GSM.09.02-Version 4.9.0. Courtesy of ETSI.)

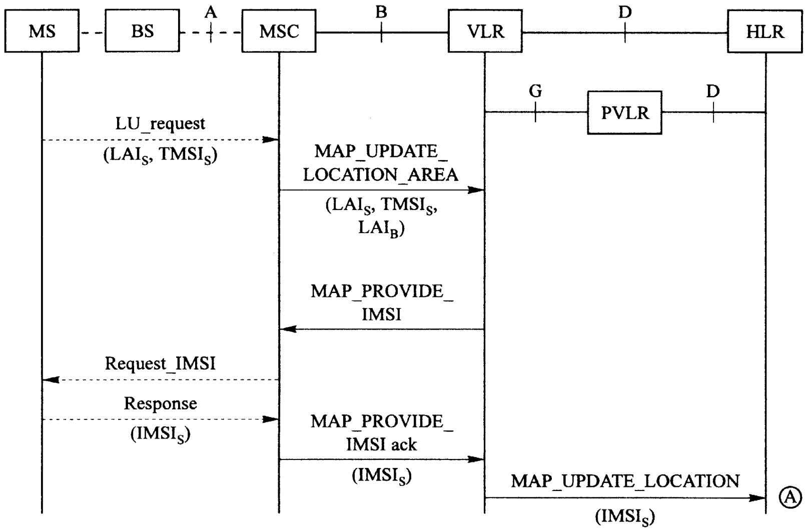


Figure 19.7-5. Initial operations for location area updating. Mobile is not known at VLR, is roaming in a foreign PLMN, and is identified by LAI_S and $TMSI_S$. (From GSM.09.02-Version 4.9.0. Courtesy of ETSI.)

In Fig. 19.7-5, MS has again identified itself by $TMSI_S$ and LAI_S. However, VLR now determines that LAI_S represents a location area in a foreign country. VLRs are not allowed to interrogate foreign VLRs. Therefore, VLR sends a provide_IMSI invoke to MSC, which then requests the IMSI from MS and returns it to VLR. Visitor location registers are allowed to interrogate foreign HLRs, and VLR derives the HLR address from IMSI and accesses HLR (A).

19.7.3 Subsequent Operations

We now explore the operations that follow the receipt of the update_location invoke by HLR (A). In Fig. 19.7-6, HLR initiates a cancel_location with PVLR, which now no longer serves MS. The address of PVLR is in the VLR field of the HLR record for the mobile. The invoke identifies the mobile by its IMSI. PVLR then deletes its MS record for the mobile.

The HLR next updates its MS record, replacing the address of PVLR with the address of VLR. It then initiates an insert_subscriber_data operation with VLR. The invoke includes the information in the MS record at HLR. VLR establishes its MS record, and VLR and HLR then acknowledge each other's invokes (B).

The operations following (B) are shown in Fig. 19.7-7. VLR first requests the AUC and the mobile to execute the verification algorithm and checks whether the results SRES are equal. We assume that this is the case. Otherwise, VLR aborts the location update.

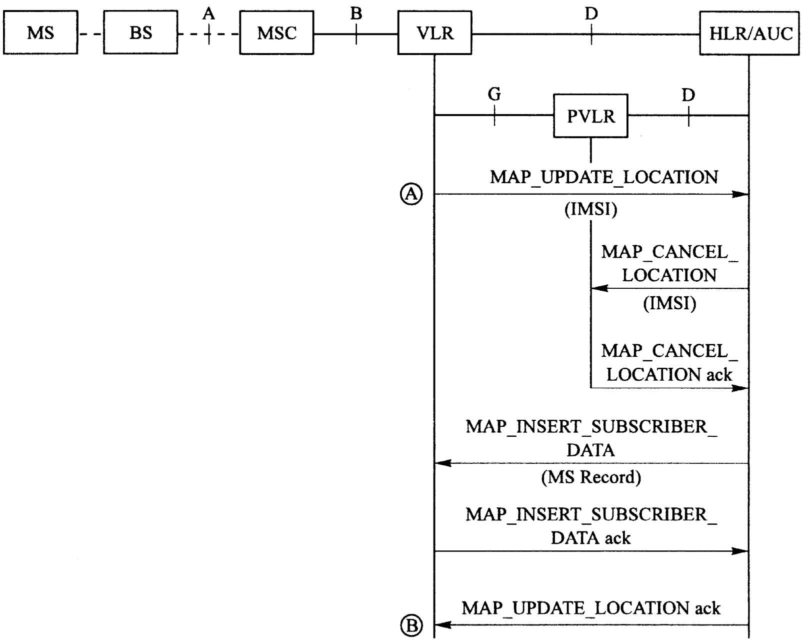


Figure 19.7-6. Subsequent operations for location area updating. (From GSM.09.02-Version 4.9.0. Courtesy of ETSI.)

If MS has ciphering (encryption) ability, VLR sends a MAP_SET_CIPHERING_MODE message, which includes Kc_A, to MSC. MSC sends a Cipher_Mode_Command, including Kc_A, to the BS serving MS. BS then generates the encryption masks for its communications with MS. BS also repeats the command (without Kc_A) to MS, which generates its masks, using its stored Kc. MS then sends a Cipher_Mode_Complete message. All subsequent information between MS and BS is encrypted.

If MS is new to VLR, the register allocates a new $TMSI_N$ and opens a forward_new_TMSI operation with MSC, which passes TMS_{IN} to MS. After receiving a confirmation from MS, MSC acknowledges the forward_ new_TMSI invoke.

Finally, VLK acknowledges the update_location invoke. This completes the updating procedure.

19.8 OPERATIONS FOR CALLS TERMINATING AT MS

19.8.1 Routing Calls to the Visited MSC

A call to a MS requires interworking between the fixed (PSTN/ISDN) and mobile (PLMN) networks. The fixed network extends the call setup to a gateway mobile

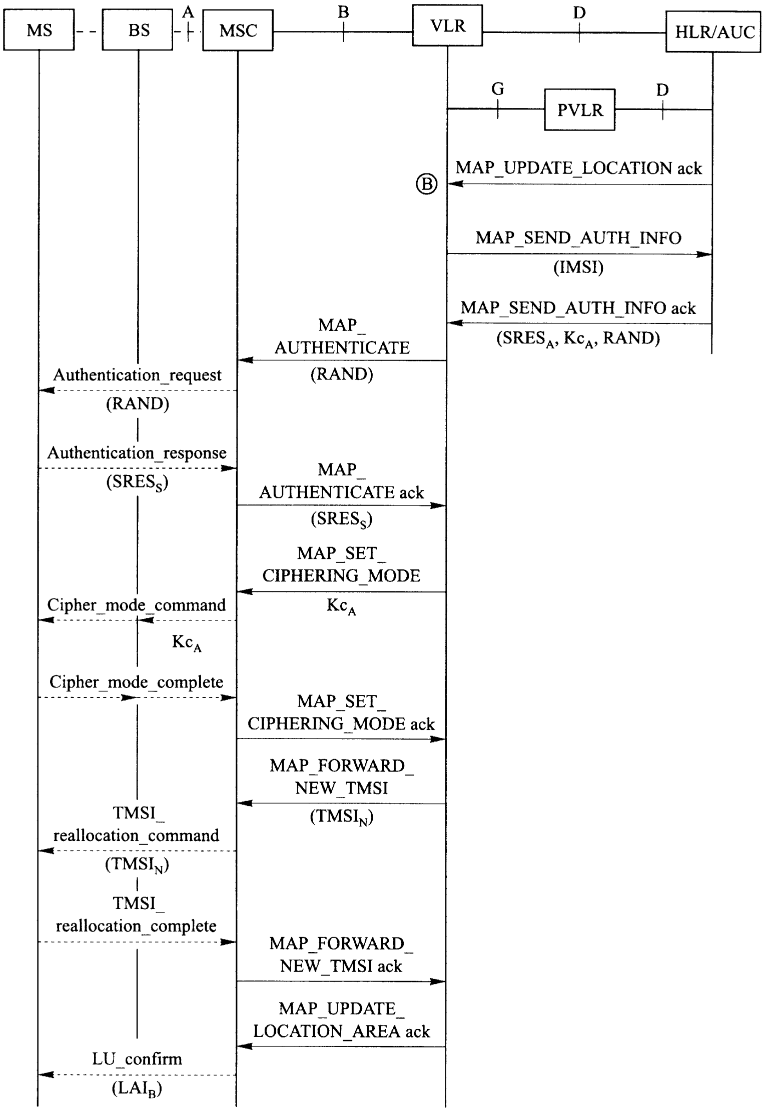


Figure 19.7-7. Final operations for location area updating. (From GSM.09.02-Version 4.9.0. Courtesy of ETSI.)

switching center (GMSC) in the home PLMN of MS, and the connection eventually reaches the MSC that is currently serving the MS.

There are a variety of possible configurations for this connection [12]. Figure 19.8-1 shows the simplest case. The calling party, the home PLMN of the called MS, and the PLMN that is currently serving MS are all in country 1. The caller

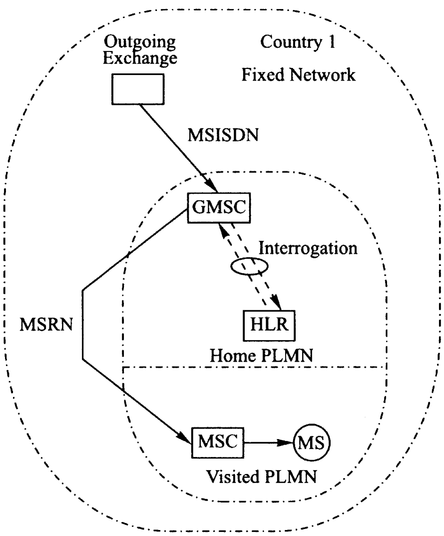


Figure 19.8-1. National connection to called mobile. (From GSM.09.02-Version 4.9.0. Courtesy of ETSI.)

has dialed the MSISDN of MS, and the fixed network uses MSISDN to set up a connection to a gateway MSC (GMSC) in the home PLMN. GMSC interrogates HLR and obtains the mobile station roaming number (MSRN), which is used to extend the connection to the MSC that is currently serving MS.

In the figure, the latter part of the connection is made in the fixed network. Some PLMNs have trunk groups between their MSCs. In that case, the connection from GMSC to MSC is made with PLMN trunks.

In Fig. 19.8-2, the calling party is in country 1, the home PLMN of MS is in country 2, and the MS is currently visiting a PLMN in country 3. The calling party has dialed the country code of country 2, followed by MSISDN, and the call setup arrives at GMSC.

GMSC interrogates HLR and obtains the country code of country 3 and the MSRN. The fixed network uses these parameters to route the call to the MSC that is serving MS.

The resulting connection is not always an economic one. Let us consider the case that countries 1, 2, and 3 are Belgium, Argentina, and France, respectively. A direct connection from Belgium to France would be more economical than a connection via Argentina.

Automatic rerouting (Section 1.3.4) can alleviate this problem in some instances. Suppose that in Fig. 19.8-3 the trunks between international switching centers ISC-A and ISC-B, and between ISC-B and GMSC, have SS7 signaling, that ISC-A has rerouting capability, and that GMSC can determine—from the calling party number received from ISC-A and the MSRN received from HLR—that the originating and terminating countries are Belgium and France. It then sends a Backward Unsuccessful message (BUM) to ISC-B. The message includes the country

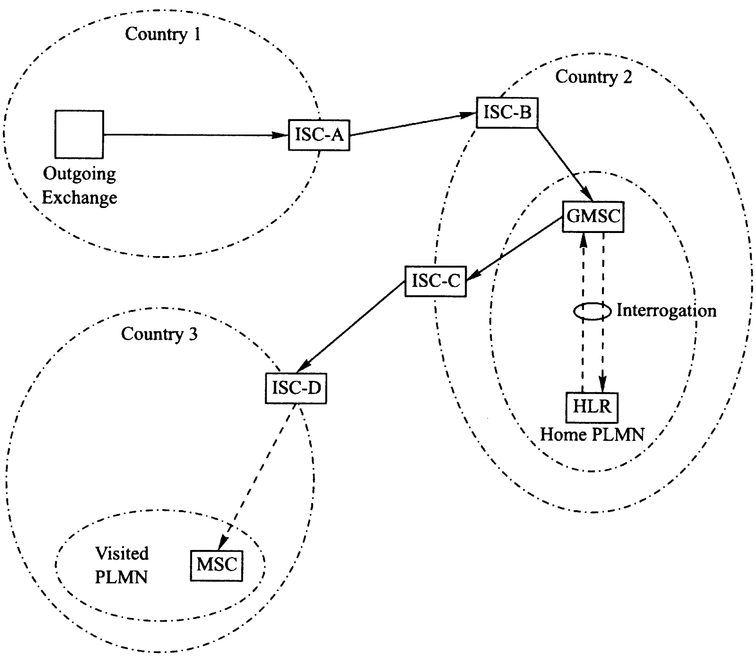


Figure 19.8-2. International connection to called mobile. (From GSM.09.02-Version 4.9.0. Courtesy of ETSI.)

code of country 3 and MSRN as the redirecting address. ISC-B repeats the message to ISC-A. This exchange then drops the forward connection and sets up a connection to ISC-D.

19.8.2 Operations for Terminating Calls

This section lists the operations involved in handling calls to a MS and includes the most important parameters in the invokes and acknowledgments [9].

MAP_SEND_ROUTING_INFORMATION. This operation is initiated by a gateway MSC in the home PLMN of MS, which has received a call for MS from the fixed network. The send_routing_information invoke is a request to the HLR of MS to provide a MSRN for the call. The called MS is identified by its MSISDN. HLR includes MSRN—and possibly the country code of the destination country—in its acknowledgment.

MAP_PROVIDE_ROAMING_NUMBER. Roaming numbers are allocated by the VLR associated with the MSC that is currently serving MS. HLR has the address of this VLR in its MS record.

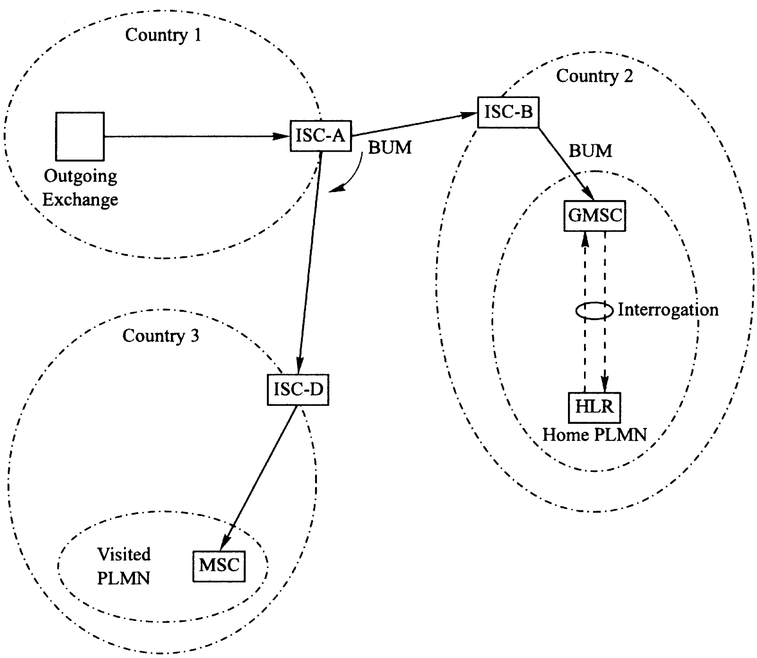


Figure 19.8-3. International connection with automatic rerouting. (From GSM.09.02-Version 4.9.0. Courtesy of ETSI.)

The HLR initiates a provide_roaming_number operation with VLR, identifying the MS by its IMSI. VLR includes MSRN—and possibly the country code of the destination country—in its acknowledgment.

MAP_SEND_INFO_FOR_INCOMING_CALL. The operation is initiated by a MSC that has received an incoming call to MS, which is identified by MSRN. It requests VLR to execute operations needed for handling the incoming call.

MAP_PAGE. The invoke is sent by VLR and requests MSC to page the MS. It includes the TMSI and LAI (location area identity) of MS.

MAP_PROCESS_ACCESS_REQUEST. The operation is invoked by MSC, after it has received a page_response from MS. It requests VLR to execute those operations that are necessary to provide access by MS to the mobile network.

VLR may decide to execute one or more of the following operations: provide_IMSI, authentication, set_ciphering_mode, and forward_new_TMSI, and then acknowledges the invoke.

MAP_COMPLETE_CALL. The invoke for this operation is sent by VLR and includes the terminating part of the MS service profile. MSC needs this information when processing the call to MS.

19.8.3 Terminating Call Procedures

In Fig. 19.8-4, the setup of a connection for a call to a MS served by MSC has reached a GMSC in the home PLMN of MS. GMSC has received the MSISDN of MS as the called party address, during interexchange signaling for the incoming trunk.

The send_routing_information invoke requests HLR to provide the MSRN for the MS. HLR derives the IMSI of MS from its MSISDN and includes it in its provide_roaming_number invoke to VLR, which is currently serving MS. VLR allocates a MSRN and enters it in its MS record. MSRN is included in the acknowledgments by VLR and HLR and reaches GMSC. GMSC now has a new called number for the call and seizes a trunk to (or toward) MSC.

In Fig. 19.8-5, the call setup has arrived at MSC. MSC initiates a send_info_for_incoming_call transaction with its associated VLR. This operation triggers a series of operations. In the first place, VLR uses MSRN to find its record of MS and retrieves TMSI and LAI. VLR then sends a page invoke to MSC that includes these parameters, and MSC broadcasts a page_request in the cells belonging to LAI.

On receipt of the page_response, MSC initiates a process_access_request operation with VLR. This register then may elect to execute one or more of the following

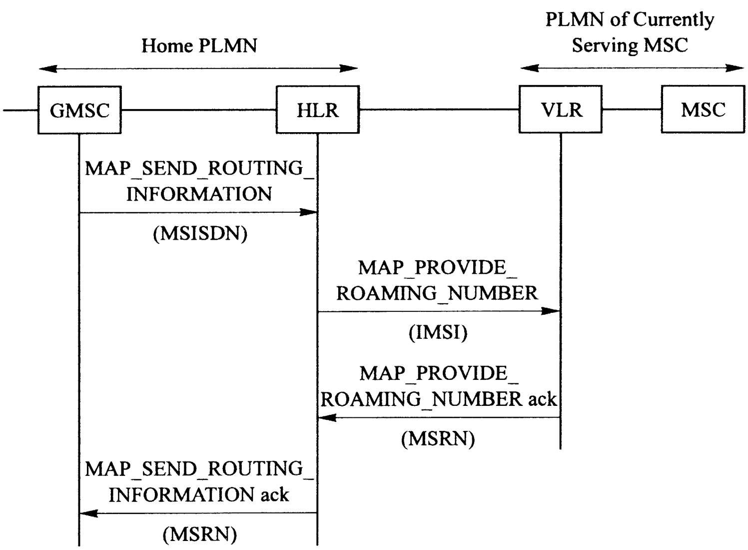


Figure 19.8-4. Obtaining a mobile station roaming number. (From GSM.09.02-Version 4.9.0. Courtesy of ETSI.)

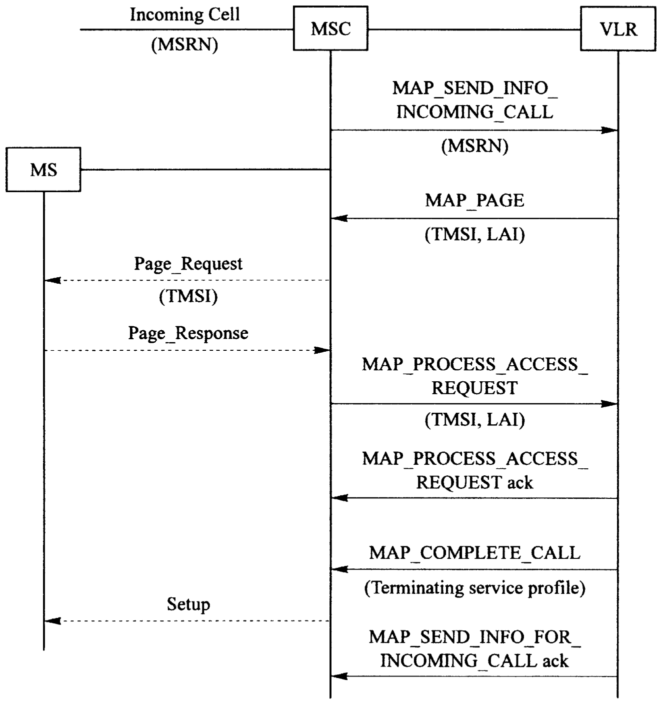


Figure 19.8-5. Operations for terminating calls. (From GSM.09.02-Version 4.9.0. Courtesy of ETSI.)

operations: authenticate, set_ciphering_mode, and forward_new_TMSI, which have been described in Section 19.7.3. The operations are not shown in the figure.

Having successfully accomplished this, VLR sends a complete_call invoke, which includes the terminating parameters of the MS service profile. MSC then starts processing the call, and VLR ends the send_info_for_incoming_call operation with an acknowledgment.

19.9 OPERATIONS AND PROCEDURES FOR ORIGINATING CALLS

19.9.1 Operations

We first explore the operations involved in handling calls originated by a MS and include the most important parameters in the invokes and acknowledgments [9].

MAP_PROCESS_ACCESS_REQUEST. The operation is invoked by a MSC that has received a CM_service_request from a MS (Section 12.9.2). It requests VLR to execute those operations that are necessary to give MS access to the mobile network.

The VLR may then decide to execute one or more of the following operations: provide_IMSI, MS authentication, set_ciphering_mode, and forward_new_TMSI. It then acknowledges the invoke.

MAP_SEND_INFO_FOR_QUTGOING_CALL. This operation is invoked by MSC and requests VLR to send information needed to process the call.

MAP_COMPLETE_CALL. The invoke for this operation is sent by VLR and includes the originating part of the MS service profile.

19.9.2 Originating Call Procedures

The procedure for handling MS-originated calls is shown in Fig. 19.9-1. On receipt of a CM_service_request, MSC initiates a process_access_request to VLR. The VLR may then elect to execute the MS authentication operations and, if MS has ciphering capability, the set_ciphering_mode operation (see Fig. 19.7-7). After completing these operations (which are not shown in Fig. 19.9-1), VLR acknowledges the process_access_request.

MSC then sends a CM_service_request_accepted message to MS. On receipt of a setup message from MS, MSC initiates a send_information_for_outgoing_call operation with VLR. VLR returns a complete_call invoke, which includes the

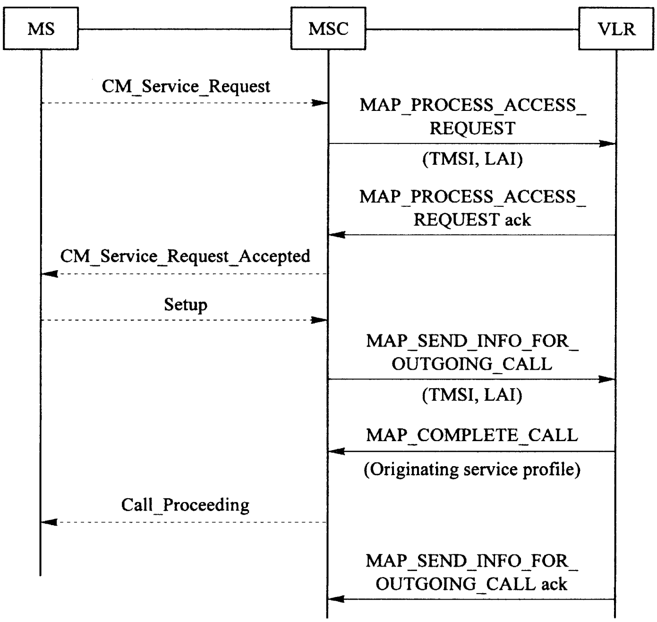


Figure 19.9-1. Operations for originating calls. (From GSM.09.02-Version 4.9.0. Courtesy of ETSI.)

origination-related parameters of the MS service profile, and then acknowledges the invoke of MSC.

MSC now has sufficient information to start the call processing and sends a call_proceeding message to MS.

19.10 ACRONYMS

AC	Area code
AIN	Advanced intelligent network
ASE	Application service element
AUC	Authentication center
AUTHR	Result of the execution of CAVE
AUTHRQST	Authentication request
BCD	Binary coded decimal
BS	Base station
CAVE	Cellular authentication and voice encryption algorithm
CDA	Called party address
CHAN	Channel
CMN	Cellular mobile network
DMAC	Digital mobile attenuation code
DTX	Discontinuous transmission
DVCC	Digital verification color code
EC	Exchange code
EIA	Electronic Industries Alliance
EIR	Equipment identity register
ETSI	European Telecommunications Standards Institute
FACDIR	Facilities directive
FACREL	Facilities release
GMSC	Gateway MSC
GSM	Global System for Mobile Communications
GT	Global title
GTA	Global title address
HLR	Home location register
HOMEAS	Handoff measurement
IMEI	International mobile equipment identity
IMSI	International mobile station identity
IS-54	Interim standard #54
ISC	International switching center

ISDN	Integrated Services Digital Network
ITU	International Telecommunication Union
ITU-T	Telecommunication Standardization Section of ITU
Ki	Authentication key
Kc	Cipher key
LAC	Location area code
LAI	Location area identity
LOCREQ	Location request
LU	Location update
MAP	Mobile Application Part
MCC	Mobile country code
MIN	Mobile identification number (ten digits)
MNC	Mobile network code
MOCH	Mobile on channel
MS	Mobile station
MSC	Mobile switching center
MSCID	Identity of MSC or VLR
MSIN	Mobile station identification number
MSISDN	Mobile station ISDN number
MSN	Mobile serial number
MSRN	Mobile station roaming number
PC_SSN	Point code and subsystem number
PLMN	Public land mobile network
PSTN	Public switched telecommunication network
PVLR	Previous visitor location register
RAND	Random number
REGCANC	Registration cancellation
REGNOT	Registration notification
RF	Radiofrequency
ROUTREQ	Routing request
SCCP	Signaling Connection Control Part
SID	System identification (15 bits)
SRES	Signed result
SSD_A	Shared secret data for authentication
SSD_B	Shared secret data for voice privacy
SSN	Subsystem number
SS7	Signaling System No.7
SWID	Switch identification
SWNO	Switch number

TCAP	Transaction Capabilities Application Part
TDMA	Time-division multiple access
TG	Trunk group
TIA	Telecommunications Industry Association
TLDN	Temporary local directory number
TMSI	Temporary mobile station identity
TN	Translation type
VC	Voice channel
VLR	Visitor location register
VMAC	Mobile attenuation code on voice channel
VPMASK	Voice privacy mask

19.11 REFERENCES

1. *Cellular Radiotelecommunications Intersystem Operations*, TIA/EIA-41 D, Telecommunications Industries Association, Washington, DC, 2002.
2. *Cellular Radiotelecommunications Intersystem Operations: Signalling Protocols*, EIA/TIA/IS-41.5 C, Electronic Industries Association, Washington, DC, 1996. [Now included in TIA/EIA-41 D]
3. *Cellular Radiotelecommunications Intersystem Operations: Automatic Roaming*, EIA/TIA/IS-41.3 C, Electronic Industries Association, Washington, DC, 1996. [Now included in TIA/EIA-41 D]
4. *Cellular Radiotelecommunications Intersystem Operations: Authentication, Signaling Message Encryption, and Voice Privacy*, Telecommunications Systems Bulletin TSB51, Telecommunications Industries Association, Washington, DC, 1993.
5. *Cellular Radiotelecommunications Intersystem Operations: Intersystem Handoff*, EIA/TIA/IS-41.2 C, Electronic Industries Association, Washington, DC, 1996. [Now included in TIA/EIA-41 D]
6. *General Aspects of Public Land Mobile Networks*, Rec. Q.1001, ITU-T, Geneva, 1989.
7. *The International Public Telecommunication Numbering Plan*, Rec. E.164, ITU-T, Geneva, 1997.
8. *The International Identification Plan for Mobile Terminals and Mobile Users*, Rec. E.212, ITU-T, Geneva, 2004.
9. *European Digital Telecommunication System (Phase 2), Mobile Application Part (MAP)*, **GSM 09.02**, European Telecommunications Standards Institute, Sophia Antipolis, France, 1994.
10. *Structure of the Land Mobile Global Title for the Signaling Connection Control Part (SCCP)*, Rec. E.214, ITU-T, Geneva, 1988.
11. *Mobile Application Part*, Rec. Q.1051, ITU-T, Geneva, 1988.
12. *General Signaling Requirements on Interworking Between ISDN or PSTN and the PLMN*, Rec. Q.1031, ITU-T, Geneva.

20

INTRODUCTION TO PACKET NETWORKS AND VoIP

Most of the previous chapters deal with circuit-switched telecommunication networks (such as the public switched circuit network—SCN), where an end-to-end connection is set up at the start of a call and released when the call ends (see beginning of Chapter 1). This chapter introduces signaling for voice connections and transmission of speech in a different type of network, namely, digital data networks that use packet communication. While circuit-switched networks were developed and optimized for voice communication, *packet networks* were developed and optimized for data communication, that is, communication between computers or *hosts* (in the rest of the chapter we will use the two terms interchangeably). Packet networks are a relatively recent development, and they are growing more rapidly than circuit-switched networks because of the large number of new computer applications in the business, government, and academic worlds that rely on host-to-host communication.

20.1 PACKET-BASED COMMUNICATION

This section presents a brief introduction to familiarize the reader with basic concepts that will be helpful later in the chapter. Readers who are interested in a more detailed description of this topic are referred to [1,2].

20.1.1 Data versus Voice

Data communication covers a range of applications (e.g., file transfer, e-mail service, message service, time-sharing, and interactive computer sessions) with characteristics that differentiate it from voice communication, namely,

Signaling in Telecommunication Networks, Second Edition, by John G. van Bosse and Fabrizio U. Devetak

- Burstiness
- Greatly variable bit rates, depending on application and/or hardware
- Tolerance of latency (delay) and jitter (variation in delay)
- Intolerance of errors in the received data
- Mostly one-way communication

Voice communication, on the other hand:

- Generates data at a steady rate for long intervals (*streaming*)
- Is very sensitive to latency and jitter (*real-time* application)
- Is more tolerant of errors in the received data
- Is mainly a two-way form of communication

Circuit-switched networks are not a cost-effective medium for data communication. Such networks are based on 64-kb/s channels, each having a preassigned temporal position—a time slot—on a transmission link (time-division multiplexing—TDM, Section 1.5). Channels are connected into a dedicated path between sender and receiver for the duration of the communication process, regardless of the availability of data to send. When such networks are used for data, valuable transmission capacity goes unused between bursts of data (e.g., during web browsing), during one-way data flows, or when the bit rate of the source is lower than the bit rate of the data link. In packet networks, on the other hand, data flows are not assigned to fixed time slots. When a host has some data to send, it segments it into relatively small pieces and forms them into *packets* by adding a *header* with addressing and sequencing information. That allows a packet to be routed independently, regardless of its position on a transmission link, and the original data to be reassembled by the recipient. A host sends a packet only when data is available, making transmission resources available to other hosts when it has nothing to send. As a result, packet networks have much higher bandwidth efficiency than circuit-switched networks, in spite of header overhead. However, because packets from a user are not assigned fixed temporal positions on links, they must be queued at each intermediate node, waiting for their turn to be transmitted along with packets from other users (statistical multiplexing). In summary, the trade-offs for bandwidth efficiency are unpredictable latency (delay), due to queuing at each node, and possible out-of-sequence delivery because of independent routing.

Since the trade-offs are acceptable for data applications, modern data networks are all packet networks. The same trade-offs, however, for many years prevented the use of packet networks for voice communication. But recent technological progress has allowed the obstacles to be overcome—as discussed in Section 20.3.

Packets are relatively small units (a typical size is about 1500 bytes), so a large piece of data must be segmented and sent as multiple packets. The reason for the size limitation is that larger pieces of data could "hog" a data link, causing unacceptable delays to smaller units queued behind them. Furthermore, if even a small portion of a large unsegmented unit becomes corrupted during transmission, the whole unit, rather than just the corrupted part, would have to be resent. Segmentation thus allows both fair sharing of resources and faster recovery from errors. Two approaches to packet size exist: the classic TCP/IP approach—discussed in this chapter—which uses longer, variable-length packets, and the ATM approach—discussed in Chapter 22—which uses shorter, fixed-length packets called *cells*.

20.1.2 Packet Networks

Packet networks can be viewed both as functional and as administrative entities. Functionally, a packet network is a collection of transmission, switching, and routing resources that interconnect hosts and use the same or compatible technology and protocols. Administratively, a packet network is an entity under the control of a single organization, such as a telecom operator or an enterprise. The distinction is useful because multiple independently managed networks may interconnect to form a larger functional network and, conversely, multiple functional networks may be interconnected and managed as one administrative entity.

A packet network consists of packet *endpoints* (hosts) and *intermediate nodes*, interconnected by digital data links. Intermediate nodes can be of two types: *switches* and *routers* (discussed later in this section). In the OSI model (also discussed later in this section), intermediate nodes are called *intermediate systems* (ISs) and endpoints are called *end systems* (ESs), but that terminology is not widely used.

A valuable characteristic of packet networks, beyond bandwidth efficiency, is that they can be self-routing and resilient: traffic can be rerouted automatically as intermediate nodes go in and out of service, without any manual reconfiguration of the network. That is made possible by *routing protocols* (see Section 22.6.1).

LANs and WANs. These are the most widely deployed types of packet networks. They are called *physical networks*, because they are identified by physical aspects such as topology, data link protocols, and hardware interfaces.

Local area networks (LANs) are packet networks where all hosts are in the same building, or in a few nearby buildings, and are interconnected by short-range transmission facilities with standard, compatible interfaces. Originally, LANs used a common bus architecture, where hosts accessed a shared transmission medium with the help of *multiple access* and *collision avoidance* protocols such as Ethernet and Token Ring (Fig. 20.1-1). In modern LANs the common bus approach (still found in the form called *hub*) has been largely superseded by packet switches. Packet switches switch packets between LAN ports based on physical addresses, eliminating collisions and bandwidth-sharing limitations.

Wide area networks (WANs) are formed by interconnecting remotely located LANs with long-distance data links. WAN data links (see Fig. 20.1-2) are point-to-point facilities, often leased from telecommunication companies and connected at each end to routers. A WAN is managed as a single administrative entity.

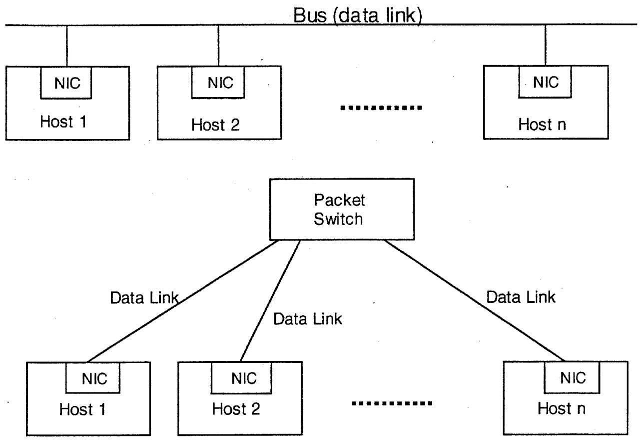


Figure 20.1-1 Local Area Network—common bus and packet switch.

Historically, LAN and WAN equipment supplied by different manufacturers used proprietary hardware and software. That made the interconnection of data networks from different suppliers difficult if not impossible. The introduction of standard network interfaces and protocols such as Ethernet [3], ATM, and TCP/IP made interconnections possible on a worldwide basis.

Internet. A packet network often needs to be connected to other packet networks in order to access a wider pool of resources. That connection requires special provisions, either because of incompatible technologies or because of separate administrative domains. Multiple individual networks connected to each other form an *internetwork* or an *internet* (with a lowercase i). The Internet (with an uppercase I) is the best-known example of an internet and consists of the interconnection of a vast number of packet networks on a worldwide basis, using the TCP/IP protocol suite for communication. The World Wide Web (WWW) is a subset of the Internet, consisting of servers that use the HTTP protocol [4]. An internet appears to a host as

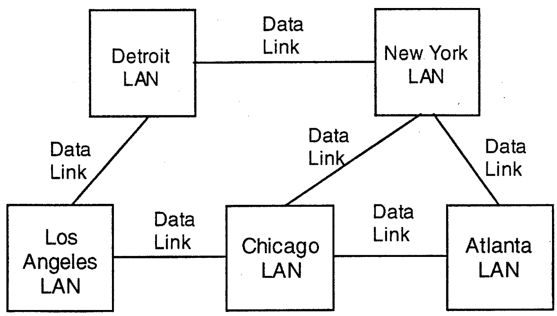


Figure 20.1-2 Wide Area Network.

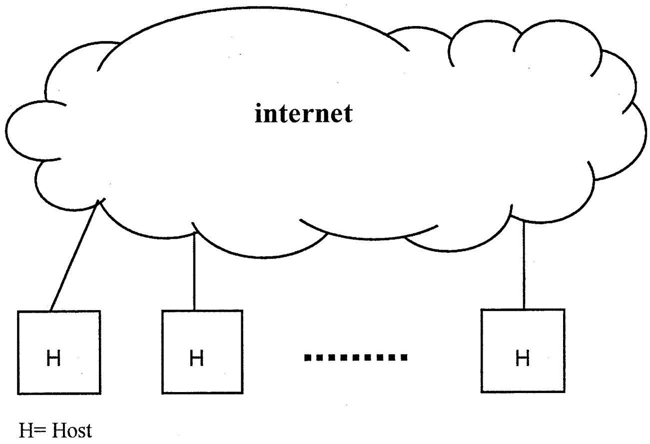


Figure 20.1-3 An internet as viewed by hosts.

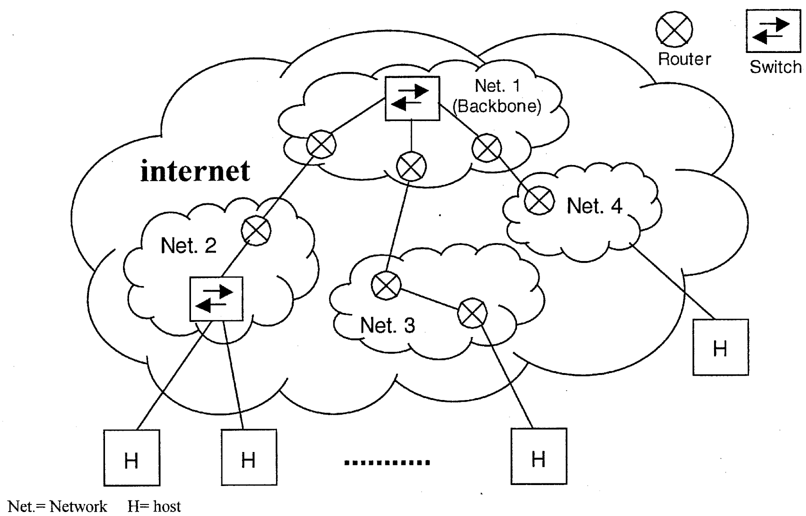


Figure 20.1-4 Internal structure of an internet.

a single packet network (Fig. 20.1-3); however, it consists of LANs, WANs, and servers interconnected by routers (Fig. 20.1-4). An internet usually has one *backbone* network that connects all other networks together. The backbone is often owned and operated by a commercial carrier and may not have any hosts connected to it directly. The success of internets is due to the fact that routers and internet protocols were developed in such a way that they required only limited hardware and software modifications to already installed "physical" networks.

In this chapter, the term "packet network" may be used to mean an individual packet network or multiple interconnected packet networks (an internet), with the distinction made obvious by the context. Also, we refer to transmission resources as data links, both in the case of point-to-point links, such as leased lines or ATM virtual circuits, and in the case of the point-to-multipoint multiple-access media found in local area networks.

Switches and Routers. Switches provide connectivity internal to a LAN and can also connect two closely located LANs provided they use the same protocol (*bridging*). Routers provide both internal and external network connectivity: internally they connect LANs to form WANs; externally they connect WANs or administratively separate networks to form internets. Routers provide protocol conversion (interworking) between incoming and outgoing links and use standard protocols to communicate with each other.

Switches and routers use *store and forward* techniques, queuing packets in interface buffers before transmission. The management of queues is a fundamental aspect of packet networks, in which queuing delays replace the call setup delay of circuit-switched networks.

Packet networks are conceptually similar to the common-channel signaling network described in Chapters 5, 7, and 8. SS7 messages are essentially packets. Signal transfer points (SCPs) are functionally similar to routers and the MTP part of the SS7 protocol is divided into levels similar to the layers of packet network protocols discussed in the next section.

20.1.3 General Principles of Packet-Based Communication

Layered Reference Models. Protocols for packet-based communication are described in terms of a hierarchy of layers, called a *reference model*, whose purpose is to provide a framework for modular design and implementation. Each layer can be viewed as a logical function that provides a service to the layer above it and communicates with the peer layer function at the other end of the connection. That enables the definition of standard interfaces between layers, which in turn allow independent development or modification of each layer's functions without affecting other layers. Two reference models are most widely used:

- The four-layer TCP/IP model, also known as the DoD (Department of Defense) or the Internet model
- The seven-layer OSI (open system interconnection) model, specified by the International Standards Organization (ISO)

The TCP/IP model was created first but is not as widely used as the OSI model, so in this chapter we follow the latter. Table 20.1-1 shows how the layers of each model map (approximately) into each other.

Layer 1 (physical layer). This layer is concerned with the electrical or optical encoding of bits and other physical aspects of the connection between the sending and receiving end of a data link (voltages, bit rates, pins, etc.).

TABLE 20.1-1 Protocol Layered Models

OSI Model	TCP/IP Model
7—Application 6—Presentation 5—Session	Process
4—Transport	Host-to-host
3—Network	Internet
2—Data link 1—Physical	Network access

Layer 2 (data link layer). This layer is concerned with the addressing scheme on the data link and with data link management. Layer 2 addresses, called physical or MAC (medium access control) addresses, are used locally on the data link or LAN, where a node or end point is connected. In a LAN they are usually “burned-in” by the manufacturer of the network interface card (NIC) of each node/endpoint. An example of a MAC address is the 48-bit Ethernet address. Link management deals with the establishment and management of data link connections, as done by protocols like Ethernet, Token Ring, PPP, and HDLC.

Layer 3 (network layer). This layer is concerned with logical network addressing and with routing between LANs, WANs, and packet networks in general. Logical addresses, also called *network addresses*, are not directly linked to physical equipment (like MAC addresses) and are configured by the network operator. They identify hosts in an internet.

Layer 4 (transport layer). This layer is concerned with host-to-host connections—their level of reliability, multiplexing, segmentation, reassembly, and flow control.

Different levels of reliability are supported by connectionless or connection-oriented communication (see later in this section).

Multiplexing allows different applications in the same host to communicate with peer applications in other hosts, and the same application in a host to communicate, in time-shared fashion, with multiple hosts. Multiplexing is achieved via *transport service access points* (TSAPs), an address mechanism that identifies an application and its instances. The L3 *logical address* and the L4 TSAP are usually paired together to form the *transport address*.

Segmentation and reassembly breaks the source data into smaller pieces that can fit into packets and reassembles them into the original data at the other end.

Flow control ensures that data is sent at a rate that can be handled by the network and by the far end without loss of information.

Layer 5 (session layer). This layer is concerned with the management of sessions between hosts. A session is an instance of communication between hosts that may include separate but interrelated data transfers as, for example, in an interactive financial transaction.

Layer 6 (presentation layer). This layer is concerned with the way data is presented to the application layer, such as encoding schemes (e.g., ASCII or binary), compression, and encryption.

Layer 7 (application layer). This layer is concerned with overall management and control of the communication process that transfers data between hosts. This layer interfaces directly with the user, be that a human or a software program. Examples of Layer 7 protocols are HTTP (Hypertext Transfer Protocol), e-mail programs, and the signaling protocols for VoIP (discussed in Section 20.3 and in the next chapter).

Layers 1 through 4 are concerned with the transport of data between hosts. Layers 5 through 7 are concerned with functions internal to hosts and play their role before a packet is sent and/or after it has been delivered.

It needs to be pointed out that the OSI model is used mainly as a reference framework and that implementations often skip layers, merge layers, or split layers into sublayers. Layers 5 and 6, in particular, are often skipped and/or their functions are merged into Layer 7.

Layers and sublayers generally correspond to separate communication protocols, which, when describing a communication process, are depicted on top of each other to form a *protocol stack*.

Types of Packet Communication. The transport layer (L4) supports two types of communication:

- Connectionless communication
- Connection-oriented communication

With connectionless communication each packet sent by a host is treated as an independent entity, with no feedback from the destination and no automatic retransmission in case of packet loss.

With connection-oriented communication the source and destination hosts establish a mutual agreement to communicate before data-carrying packets can be exchanged, thus creating a logical connection. The logical connection is maintained for the duration of the packet flow and released at the end of it. Establishment and release of the (logical) connection is done by exchanging packets that carry control information. It should be noted that, even with connection-oriented communication, no resources or paths are booked or reserved between source and destination; source and destination are the only entities aware of the connection. Intermediate nodes, which operate only up to Layer 3, are not aware of connections, which are Layer 4 functions. A similar distinction between connectionless and connection-oriented is found in common-channel signaling with the SCCP protocol (Chapter 15).

Reliability of Packet Communication. Some applications require reliable delivery of data, while others do not. Connectionless transport is inherently unreliable, since there is no way for the transport layer to ensure that packets are delivered successfully. Connection-oriented transport, on the other hand, can be made reliable by procedures that ensure delivery of packets, such as acknowledgment of receipt and retransmission of lost or corrupted packets. Reliable delivery is not an exclusive or necessary function of connection-oriented transport, although it is provided in most cases. Reliable delivery is also possible with unreliable transport, but in that case the burden falls on the application (Layer 7). Applications can be designed to detect data delivery errors and to request retransmission of lost or corrupted packets, regardless of the type of L4 protocol. In that case, however, every application would have to include a reliable delivery function and that would likely result in a multitude of dissimilar error-control procedures. A more efficient alternative is to incorporate reliability in a Layer 4 connection-oriented protocol that can serve all applications.

20.1.4 Packet-Based Communication

The basic operation of packet communication can be explained by looking at Fig. 20.1-5, which shows how the OSI layered model translates into the internal structure of a packet. Each layer corresponds to a logical entity that processes information received from an adjacent layer and then passes the result to the next layer: from a higher to a lower layer when sending, from a lower to a higher layer when receiving.

Let us start with a user who wants to send a unit of data to a distant host. The user invokes a communication-oriented application (Layer 7), which acquires the data and starts the communication process by adding a header to the data and passing the result to Layer 6. Each layer adds a header (and often at Layer 2 a trailer as well) to the data it receives from the higher layer, in a process called *encapsulation*. Each layer passes the partially encapsulated data to the next layer, until it reaches Layer 1, where the logical bits of the fully encapsulated data are encoded into physical signals and transmitted to the far end. At some stage of this process (Layer 4), the original unit of data is broken into segments if too large to fit into one packet, but enough information is put in the header of the layer doing the segmentation to allow the receiving end to reassemble it.

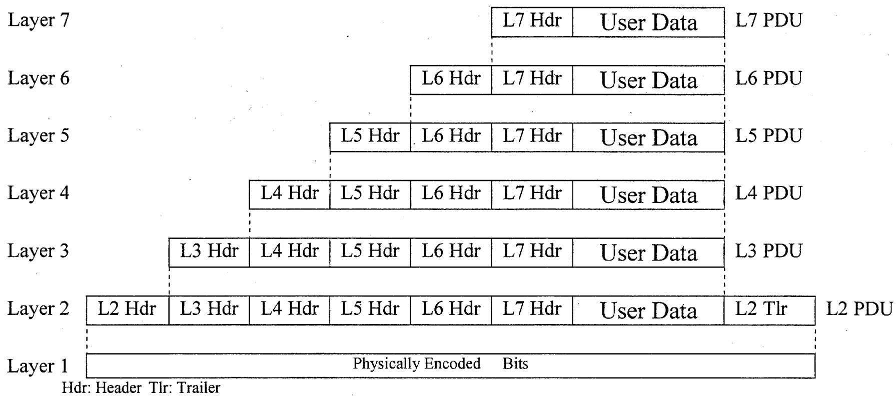


Figure 20.1-5 Encapsulation and de-encapsulation process.

At the far end the fully encapsulated data is received by Layer 1, where the physical signals are decoded into logical bits and the process is reversed: each layer strips the outermost header (and trailer if present) and processes it before passing the partially deencapsulated data to the next layer; the process continues until the original data from the user is reconstituted at Layer 7. Throughout the encapsulation and deencapsulation process, the original data from the user remains unchanged, except for Layer 6, where format conversion may take place.

Protocol Data Units (PDUs). The partially encapsulated data that is passed from one layer to the next is called a *protocol data unit* (PDU). A PDU that is passed downward from layer n to layer $n - 1$ is called a Layer n PDU. Conversely, a PDU that is passed upward from layer $n - 1$ to layer n is a Layer $n-1$ PDU. As shown in Fig. 20.1-5, a Layer n PDU consists of the Layer n header followed by the headers for the $<n$ layers and by the user data.

A Layer n PDU, once encapsulated by Layer $n - 1$, is referred to as a Layer n *service data unit* (SDU). For example, after Layer 3 adds its header, it sends a L3 PDU to Layer 2. Layer 2 then adds its own header and sends Layer 1 a L2 PDU containing a L3 SDU.

Figure 20.1-6 shows how each layer of the reference model communicates with the peer layer at the other end of the connection by exchanging PDUs through the services of the lower layers. The header (and trailer) for a layer carries the control informational for peer-to-peer communication, while the SDU inside each PDU is treated as payload and left untouched.

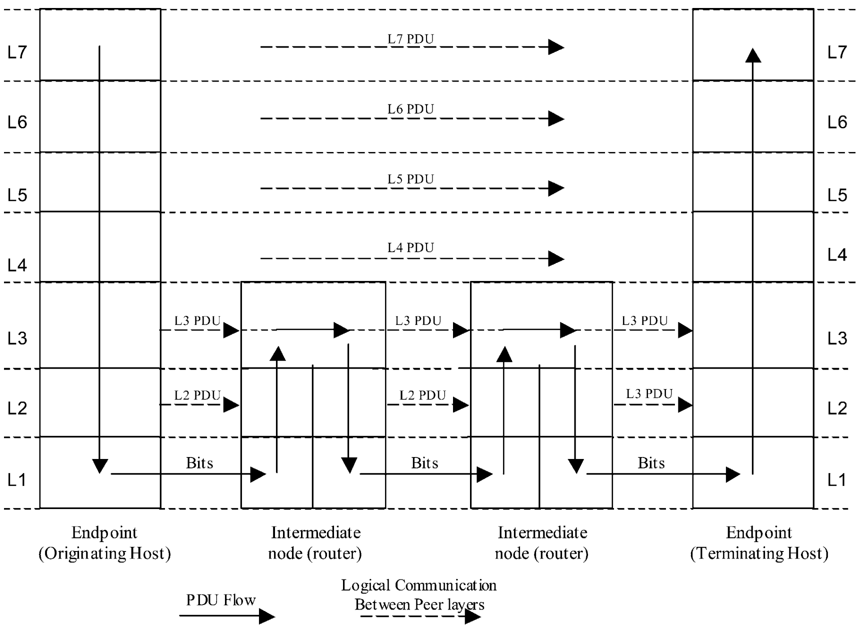


Figure 20.1-6 Peer-to-peer communication and PDU flow.

Layers in Hosts, Switches, and Routers. Not all components of a packet network perform functions corresponding to all the layers (Table 20.1-2): hosts (packet endpoints) perform functions corresponding to all seven layers, while switches and routers (intermediate nodes) perform functions corresponding to Layers 1–3 only. Specifically, switches operate at Layers 1 and 2; routers operate at Layers 1–3.

The Routing Process in Switches and Routers. This process can be summarized as path determination plus interworking between incoming and outgoing links. When a router's Layer 2 receives a L2 PDU (from the incoming link's queuing buffer), it strips the L2 header and trailer and passes the resulting packet to Layer 3 for processing. The L3 logic uses the L3 header to decide which outgoing data link to forward the packet to (*routing*) and then passes the packet, unchanged, to the L2 logic that controls the chosen outgoing data link (*forwarding*). The L2 logic adds a new L2 header and trailer to match the characteristics of the outgoing data link and then puts the reencapsulated L2 PDU into the outgoing link's queuing buffer. So, while L2 headers and trailers are replaced every time a packet goes through a router; L3 headers are processed but left unchanged. Switches only process the L2 header (without replacing it) and use it to route the packet to the appropriate outgoing port. L4 headers are ignored by switches and routers and are processed only by the originating and terminating host. The above observations are summarized in Table 20.1-2.

Note. The description of packet-based communication thus far has been based on classic routing, as done in the Internet, where all packets are treated equally. When *quality of service* (QoS) considerations come into play (e.g., when certain types of packets must be given higher priority), different schemes that deviate from what is described above may be used. For instance, a router may look into the L4 header to help with routing decisions or may use *label switching*. In label switching a short additional header (called a *label*) is prepended to the packet; the label is replaced at every node and is used to speed up routing and to facilitate QoS enforcement.

TABLE 20.1-2 Functional Aspects of the OSI Layers

Layer	Peer-to-Peer Connection	Function
Layer 7 (application)	Host-to-host	Application
Layer 6 (presentation)	Host-to-host	Application service
Layer 5 (session)	Host-to-host	Application service
Layer 4 (transport)	Host-to-host	Data transport
Layer 3 (network)	Node-to-node or host-to-node	Data transport
Layer 2 (data link)	Node-to-node or host-to-node	Data transport
Layer 1 (physical)	Node-to-node or host-to-node	Data transport

20.2 THE TCP/IP PROTOCOL SUITE

20.2.1 Overview

The TCP/IP suite is by far the most widely used set of protocols for Layer 3 and Layer 4 and is specified by the Internet Engineering Task Force (IETF—Section 2.2). It was originally based on the four-layer DoD reference model, but most current literature describes it according to the OSI model.

The main components of the TCP/IP protocol suite are IP, TCP, and UDP. The suite also contains other protocols, such as ICMP and ARP [1], which perform ancillary functions and are not discussed in this book. The protocol stack is shown in Fig. 20.2-1. That figure shows two examples of Layer 2 protocols, not part of the TCP/IP suite: a LAN protocol, like Ethernet, broken into two sublayers (MAC and LLC), and the Point-to-Point Protocol (PPP) commonly used for dial-up connections. A detailed discussion of Ethernet and PPP are found in [3,5].

In the TCP/IP suite (Fig. 20.2-2):

- Layer 4 PDUs are called *segments*.
- Layer 3 PDUs are called *packets* or *datagrams*.
- Layer 2 PDUs are called *frames*.

The term "packet" can be used both in the general sense introduced at the beginning of this chapter and in a specific sense to distinguish a L3 PDU from a L2 or L4 PDU. The context usually makes the sense clear.

Higher Layer Protocols		Higher layers
TCP	UDP	Layer 4 (Transport)
IP		Layer 3 (Network)
LLC	PPP	Layer 2 (Data Link)
MAC		
Physical Layer		Layer 1 (Physical)

Figure 20.2-1 TCP/IP protocol stack.

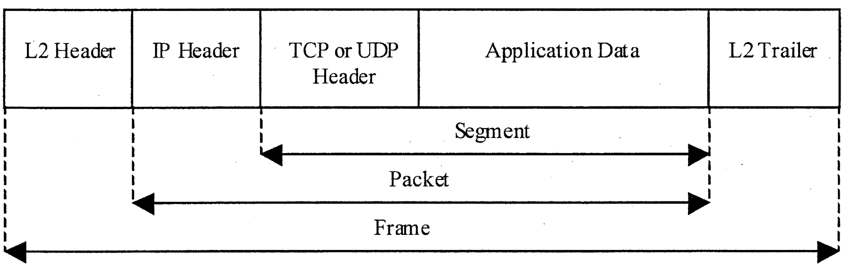


Figure 20.2-2 TCP/IP PDU.

Figure 20.2-3(*a*) shows the peer-to-peer communication and Fig. 20.2-3(*b*) shows the PDU flow when TCP/IP is the transport mechanism and the presentation and session layers (as is often the case) are not explicitly separated from the application.

20.2.2 The Internet Protocol (IP)

The main function of IP is to provide packet delivery from a source host to a destination host, using *logical addressing* and (when necessary) *fragmentation*. Each packet is treated as an independent unit and there is no provision for guarantee of delivery [6,7]. IP is a node-to-node protocol operating at Layer 3 (network layer). The header of an IP packet is processed by routers, which select the outgoing data link based on the destination IP address. Originating and destination hosts use IP addresses (in combination with *port numbers*—see next section) to identify packet flows.

There are two versions of IP: IPv4 and IPv6. IPv4 is the classic, most widely used version, and the one discussed in this chapter. IPv6 is a newer version, created primarily to eliminate an addressing bottleneck and not yet widely deployed. Various address-saving techniques have been devised since work on IPv6 was started, which have ensured the viability of IPv4 for the time being.

The fields in the IP header (Fig. 20.2-4) are:

- *Version*. The version of the format of the header.
- *IHL* (*Internet Header Length*). The number of words (4 octets) in the IP header. The minimum value is 5.

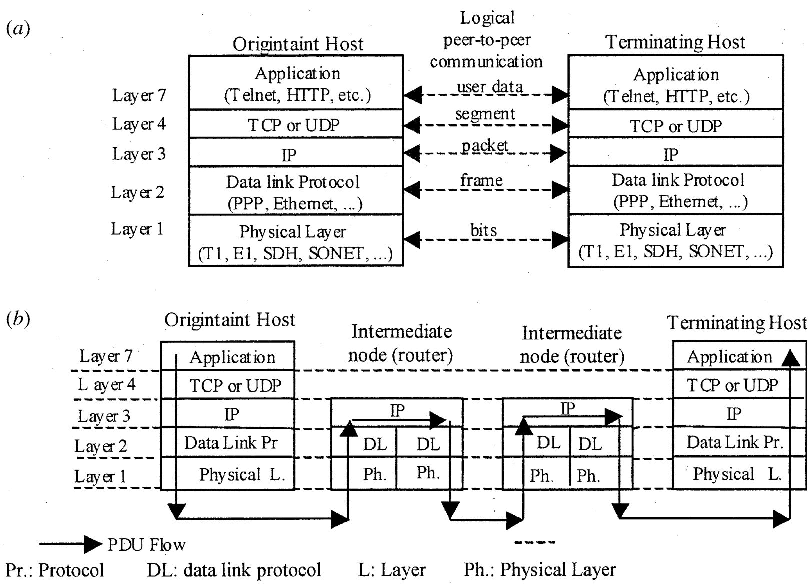


Figure 20.2-3 Peer-to-peer communication and PDU flow with TCP/IP.

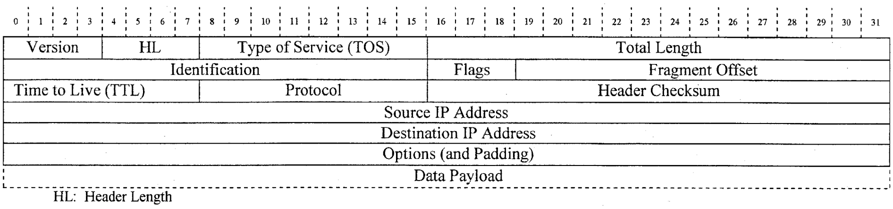


Figure 20.2-4 IP packet format.

- *TOS* (*Type of Service*). Identifies the desired quality of service and can be used by an IP network to prioritize packets (e.g., DiffServ—Section 20.3.1).
- *Total Length.* Number of octets in the packet (header plus data).
- *Identification.* Sequence number used for reassembling fragments.
- *Flags.* Indicates if fragmentation is allowed and if more fragments are coming.
- *Fragment Offset.* Indicates the position of the fragment in number of octets.
- *TTL* (*Time to Live*). Indicates the number of seconds after which the packet must be deleted. It is decreased by one every time the packet is processed by a node.
- *Protocol.* Indicates the Layer 4 (transport) protocol carried in the data portion of the packet.
- *Header Checksum.* Error-detection checksum calculated on the IP header (see Section 20.2.3 for the algorithm; IP does not use a pseudoheader).
- *Source Address and Destination Address.* The IP address of the source and the destination hosts.
- *Options.* Optional, variable-length field that carries control and testing parameters [6].

IP Addresses. IP addresses are logical addresses that identify a packet network and an individual host in that network. An IP address is functionally equivalent to an international telephone number in its ability to identify a host on a worldwide basis. IP uses a 32-bit address subdivided into 4 octets, each with a value from 0 to 255. An IP address is typically represented in *dotted decimal form*, where the decimal value of each octet is separated from the others by a dot, for example, 192.28.181.38. The leftmost portion of the address identifies a packet network and the rightmost portion identifies a host in that network. Addresses are subdivided into three *classes*, based on the boundary between the network and the host portion of the address, as shown in Table 20.2-1. Classes allow the creation of networks of different sizes, with the largest size corresponding to Class A and the smallest size to Class C. The explosion of the Internet, however, has caused the number of hosts and networks to skyrocket, leading to a tight supply of available addresses and to the need for finer granularity. The need spurred the creation of schemes for *classless addressing*, such as *subnetting* and *supernetting*, and the use of *subnet masks* to define the boundary between the network and the host portion of the address

TABLE 20.2-1 IP Address Classes

			First Octet	
Class	Network Address	Host Address	Decimal Value Range[a]	Value of Initial Bits
A	1st octet	2nd, 3rd, 4th octet	1–126	0
B	1st, 2nd octet	3rd, 4th octet	128–191	10
C	1st, 2nd, 3rd octet	4th octet	192–223	110

[a]Some values are reserved and not available to users.

[8,9]. With subnetting some of the leftmost bits of the host portion of the address (in a *classful* sense) are used to extend the network portion, freeing bits to increase the number of addressable networks. In supernetting the opposite is done, freeing bits to expand the number of hosts in a network. Another effective scheme goes by the name of *network address translation* (NAT) [10]. With NAT, a packet network uses *public addresses* from a small pool, assigning them to users when they need to access the Internet and returning them to the pool after use. *Private addresses*, assigned with no limitations, are used for intranetwork communication.

Domain Names. A normal Internet user rarely sees IP addresses thanks to a user-friendlier addressing scheme based on a hierarchical naming structure (*domain name system*—DNS [11,12]). DNS is based on *domain* and *subdomain names*, represented by alphanumeric strings, called *labels*. Labels are separated by dots, such as in www.wiley.com. A worldwide network of hosts, called DNS servers, provides translation between domain names and IP addresses. A domain name is *fully qualified* (FQDN) if it contains all the labels necessary to identify a particular host on a global basis. For efficiency, non-fully-qualified names may be used locally, when there is no ambiguity.

Fragmentation and Reassembly. This is the other basic function of IP, besides addressing. Different networks have different limits on packet size (*maximum transmission unit*—MTU), depending on the characteristics of their data links. Source hosts have no knowledge of the MTUs of the networks traversed by packets and usually limit outgoing packets to a default value of 1500 bytes. When the IP logic in a router determines that the size of a received packet exceeds the MTU of the outgoing link, it breaks the packet into smaller packets called *fragments*. The original packet is reassembled by the router where the packet exits the network with the smaller MTU. A flag bit and the parameters Identification and Offset in the IP header enable reassembly of the packet fragments.

20.2.3 The Transmission Control Protocol (TCP)

TCP is a transport layer protocol that uses the packet delivery services of IP to provide host-to-host communication [13].

The main functions of TCP are multiplexing, reliable delivery, and ordered delivery of packets. Multiplexing is done by the use of *ports*, the TCP term for what the OSI calls TSAP (Section 20.1.3). Port numbers are internal addresses by which a host (identified by an IP address) can multiplex several users accessing the same or different applications residing in it. The combination of IP address and port number (the transport address) is called s*ocket*. The source and destination sockets uniquely identify a TCP connection.

TCP segments the user data received from an application (directly or through the session layer) so it can fit into IP packets and then delivers it to the destination by communicating with the peer TCP entity in the remote host.

TCP communication between hosts is connection-oriented and provides reliable and ordered packet delivery, with retransmission of lost or corrupted packets.

The TCP header is processed by the source and destination hosts and is generally ignored by routers. The fields in the header (Fig. 20.2-5) are:

- *Source Port and Destination Port*. The port number of the source and the destination.
- *Sequence Number*. Used in sequencing octets (see three-way handshake, below).
- *Acknowledgment Number*. Identifies the next sequence number expected from the far end by the sender of the segment.
- *Data Offset*. Number of words (32 bits) in the header.
- *Reserved*. For future use, currently set to zero.
- *Control Bits*. Used in the three-way handshake (see below).
- *Window*. Number of octets that the sender will accept (from the far end) without acknowledgment.
- *Checksum*. Error-detection checksum (see below).
- *Urgent Pointer*. Identifies the length of a sequence of urgent data.
- *Options*. Variable-length field. The only defined option is the maximum segment size accepted by the sender.

The Source Port Number is usually assigned at random by the source while the Destination Port Number is usually a predefined identifier for an application,

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31

<table>
<tr><td colspan="3">Source Port</td><td>Destination Port</td></tr>
<tr><td colspan="4">Sequence Number</td></tr>
<tr><td colspan="4">Acknowledgement Number</td></tr>
<tr><td>HL</td><td>Reserved</td><td>Control Bits</td><td>Window</td></tr>
<tr><td colspan="3">Checksum</td><td>Urgent Pointer</td></tr>
<tr><td colspan="4">Options (and Padding)</td></tr>
<tr><td colspan="4">Data payload</td></tr>
</table>

HL: Header Length (Offset)

Figure 20.2-5 TCP segment format.

called a *well-known port number*. Well-known port numbers are in the range 1 to 1024, for example, 53 (DNS) or 80 (HTTP); randomly assigned port numbers are greater than 1024.

A TCP connection between two sockets can be used to send data both ways (full duplex).

TCP Connections. TCP connections use sequencing of octets and acknowledgments of receipt to check for loss of data, to request retransmission of lost or corrupted data, and to reassemble the user data at the destination. TCP is bit stream oriented so sequence numbers and acknowledgments refer to octets rather than segments.

Acknowledgments consist of segments in which the Acknowledgment Number field contains the sequence number of the next octet expected by the sender of the acknowledgment (i.e., the sequence number of the last correctly received octet plus one), and where the ACK bit in the Control Bits field is set to 1 (to indicate that the segment contains a valid acknowledgment). An acknowledgment segment can also carry user data.

The logical connection between two TCP sockets is set up by synchronizing their sequence numbers through an exchange of segments containing no user data and called a *three-way handshake* (Fig. 20.2-6). The originating host starts by sending a segment with its initial sequence number in the Sequence Number field, and with SYN = 1 in the Control Bits field. The destination host responds with a segment in which:

- The Acknowledgment Number field is set to the origination's initial sequence number plus one (e.g., if the initial sequence number was 100, the Acknowledgment Number is 101).
- The Sequence Number field is set to the destination's initial sequence number.
- The Control Bits field has ACK = 1 and SYN = 1.

The originating host then completes the handshake by sending the third segment, in which the Acknowledgement Number field is set to the destination's initial sequence number plus 1, and ACK = 1.

Once the two TCP entities are synchronized, the connection is *established* and reliable exchange of data begins. Reliable data delivery is achieved by a flow control technique called *windowing*. Windowing allows a source to send a certain number of octets without acknowledgment, based on the value of the Window field set by the destination. After the set number of octets is sent, the source stops and waits for an acknowledgment from the destination. The acknowledgment restarts the flow and also tells the source which octet to send next, thus allowing packet retransmission if needed. The TCP entity at the destination reconstructs the octet sequence and reassembles the original data. Figure 20.2-6 shows a connection with a Window value of three.

The release sequence (Fig. 20.2-6) uses the FIN (finished) bit in the Control Bits field and starts when one side sends a segment with FIN = 1. The receiving host

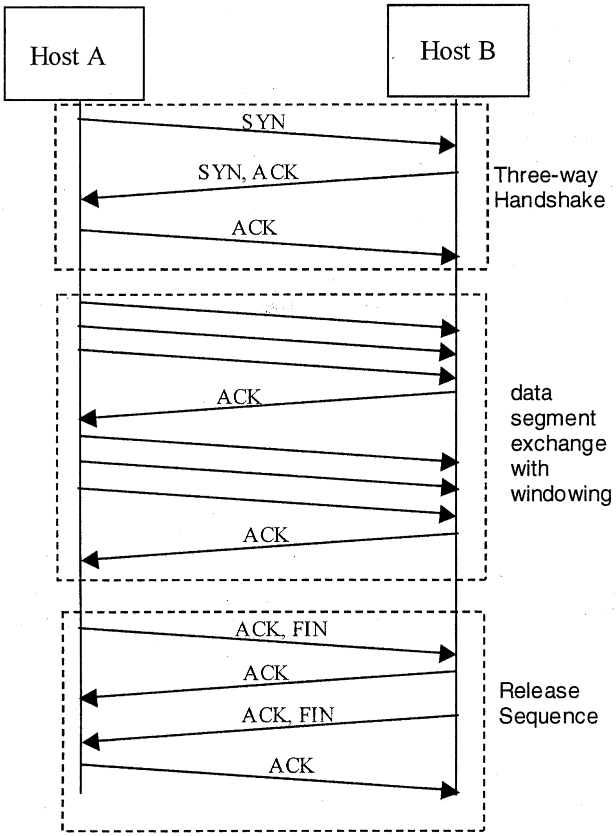


Figure 20.2-6 Example of TCP segment sequence.

acknowledges the receipt of the FIN request (with ACK = 1) while waiting for confirmation from the Layer 4 process. Once the L4 process has accepted the release, the receiving host sends the final confirmation (with ACK = 1, FIN = 1), which is acknowledged by the originator of the release (with ACK = 1) to close the sequence.

The long header and acknowledgment procedures make segment delivery relatively slow. TCP's main strength is reliable delivery, so this protocol is used primarily when loss or corruption of data is not acceptable, as with file transfer and signaling applications.

Checksum. This 16-bit parameter in the TCP header enables the L4 recipient of a segment to determine (in most cases) whether the received message is error-free. It is similar to the check bit parameter in the SS7 Message Transfer Part (Section 8.3.2). The sender calculates the Checksum by treating the segment as a sequence of 16-bit strings (padding is added as necessary). A *pseudoheader*, containing the source and destination IP addresses, is included in the calculation to protect against segment misrouting, but is not sent. The sending host performs a 16-bit ones' complement sum of pseudoheader, header, and data, and then enters the ones' complement of the result in the Checksum field. On receipt of the segment, the destination host independently computes the Checksum. If the received and computed values match, the destination host considers the message to be error-free and passes it to the application; otherwise the message is discarded.

20.2.4 The User Datagram Protocol (UDP)

UDP is the other main transport layer protocol for host-to-host communication that uses the packet delivery services of IP [14]. UDP is a connectionless protocol and does not guarantee reliable or sequential packet delivery. Its main function is multiplexing, achieved, like TCP, with port and socket.

UDP takes a unit of data received from an application and sends it to the remote host with no provisions for segmentation, acknowledgment of receipt, or retransmission of lost or corrupted segments. UDP packets are called "segments," although segmentation is not supported, and also *datagrams* (from "telegrams"), alluding to the lack of acknowledgments. Without segmentation, a unit of data sent via UDP is limited in size to the MTU of the underlying IP. Some form of segmentation or sequencing may, of course, be provided by the application or by an additional protocol (a sublayer on top of UDP).

The UDP header is processed by source and destination hosts and is generally ignored by routers. The fields in the header (Fig. 20.2-7) are:

- *Source Port and Destination Port.* Port number of the sending and of the receiving application, respectively. The Source Port is optional and when not used is set to a value of zero.
- *Length.* Number of octets in the UDP datagram. Unlike TCP, the Length field covers both header and data.
- *Checksum.* Multiple of 16 bits, padded with zeros as needed. Calculated, as for TCP (Section 20.2.3), over pseudoheader, header, and data.

Although UDP's connectionless transport is unreliable, some protection from errors is provided by the Checksum parameter, which works exactly as for TCP and allows the receiving UDP to detect most segment errors. If an error is detected, the segment is dropped and the receiving application is not informed.

0–15	16–31
Source Port	Destination Port
Length	Checksum
Data Payload	

Figure 20.2-7 UDP segment format.

The absence of acknowledgments and the shorter header make UDP segment delivery faster than TCP's. UDP is used primarily when reliable delivery is not critical, especially if fast packet transfer is important (e.g., for voice communication).

20.2.5 URI and URL

IP addresses (or the equivalent FQDNs—Section 20.2.2) uniquely identify hosts, but a host may contain multiple Internet resources (applications). An addressing scheme with finer resolution is therefore needed so a user can identify an application within a host. The IETF has defined such an addressing scheme, called URI (*uniform resource identifier*) in its more general form [15] and URL (*uniform resource locator*) [16] in its most widely used form. A URI/URL can be viewed as a user-friendlier form of a TCP/UDP socket, similar to how a domain name is a user-friendlier form of an IP address. URIs and URLs are used and processed by applications (Layer 7).

URI is a character string that identifies a resource by name or location, with components separated by "/" in case of a hierarchical structure.

URL is the subset of URI that identifies the resource by location in the network. The standard form of a URL for IP networks is

$$//\langle \text{user} \rangle : \langle \text{password} \rangle @ \langle \text{host} \rangle : \langle \text{port} \rangle / \langle \text{url-path} \rangle$$

(*Source*: RFC 1738 [16])

Not all URL parameters are mandatory. The "host" parameter is an IP address or a FQDN.

20.3 INTRODUCTION TO VoIP

One of the most recent developments in telecommunications is the use of packet technology for voice and video (collectively referred to as *media*) communication. Given the current preponderance of voice, and since TCP/IP clearly has become the leading protocol suite, *Voice over Internet Protocol* (VoIP) is the term that tends to be used when talking about the transmission of media streams in packet networks. Virtually all concepts and protocols described in this chapter apply to media communication [17–20], although voice is the key aspect.

Because media flows are transported by the same packet networks that carry data, VoIP brings about the *convergence of voice and data*. The goal of convergence is the evolution of packet networks into *next-generation networks* (NGNs). A NGN is a packet-based, open interface, transmission and switching medium that provides broadband multimedia (voice, video, data) communication for wireline and wireless users. Once NGNs become a reality, a global converged network ("the" NGN) will replace both the public switched-circuit network (Chapter 1) and public data networks.

There are two main avenues of convergence for voice and data:

1. Voice traffic and services originated and terminated in the PSTN/ISDN
2. Media traffic and services between packet-based terminals connected by a packet network or an internet.

In the first case, growth in traffic originated/terminated in the PSTN/ISDN, coupled with the unwillingness of telecom operators to invest in SCN infrastructure, leads to the use of packet infrastructure for capacity relief. In this case end-to-end transparency of ISUP-based narrowband services, as currently provided by the PSTN/ISDN, is a must, and support of multimedia and broadband services is not a requirement. This avenue of convergence is called *voice telephony over IP* (or *over ATM*, when ATM is used for transport), where the keywords *voice telephony* indicates PSTN/ISDN narrowband services.

In the second case packet terminals directly connected to packet networks need access to next-generation services based on broadband multimedia communication, in addition to basic telephony services. In this case full end-to-end transparency of PSTN/ISDN services is not a mandatory requirement since the focus is on next-generation services. This avenue of convergence is referred to as *IP telephony*.

A third case, where traffic originates in the SCN and terminates in a packet network, or vice versa, falls between the first two and can be considered a variation of the second case, since it relies on the subset of services that are supported in both domains, that is, basic telephony services.

20.3.1 Basic Concepts and Overview

VoIP is speech communication in an IP packet network where speech is transmitted and switched as a flow of packets containing digital voice samples. Achieving adequate quality of service for voice communication presents serious technological challenges due to the fact that packet networks are optimized for data communication (Section 20.1.1). To explore those challenges and how they can be overcome, we need first to consider that VoIP has two distinct functions:

1. Transport of sequences of voice samples, referred to as *media streams, media connections, bearer connections*, or the *user plane*
2. Transport of signaling messages used to establish, manage, and release media connections between communicating parties, referred to as *signaling channels*, *control channels*, or the *control plane*.

Most difficulties are found in the handling of media streams. Signaling adapts more easily to a packet environment since it shares basic characteristics with data communication.

Controlling packet delays in media streams is the main technological problem: delays in packet networks result from packetization, jitter control, and flow control. Packetization delay is caused by the time it takes to accumulate a certain number of voice samples before transmission; a typical method is to put 20 or 30 ms worth of speech samples into a frame and then ship the frame in a packet. Jitter buffers, used to control jitter, also introduce delay, but the main contributors are flow-control (queuing) buffers, which introduce unpredictable delays at each node traversed by a packet. In the extreme case of buffer overflow, packets are lost. ITU-T standards (Rec. G.114) state that end-to-end delays of up to 150 ms are fully acceptable, delays greater than 400 ms are unacceptable, and values in-between are acceptable but noticeable by users. Acceptable delays are not easy to achieve because an IP packet network does not reserve resources for the duration of a call. Each packet has to compete for resources at every intermediate node it traverses, so its delay is affected by the number of hops and by traffic patterns at each node. A burst of traffic or unusually large packets in a queue can have serious impact on delay.

The first step in addressing delay is the use of UDP as a transport protocol instead of TCP, which is too slow for streaming applications. UDP has the drawback of unreliable communication, and that is partially addressed by voice applications with algorithms such as *packet loss concealment* (PLC), which compensates for random packet loss of up to almost 5% without significant effect on voice quality. But even UDP and PLC are not sufficient, so other techniques must be used, such as (1) (over) engineering the network to ensure that even with high traffic there is enough capacity to meet QoS objectives and (2) giving priority to voice over data.

The second technique can be implemented by two methods: *differential service* or *label switching*. Differential service (DiffServ) uses bits in the TOS field of the IP header as priority flags (Fig. 20.2-4). In label switching (see bottom of Section 20.1.4), a label containing priority information is prepended to the header of a packet and, due to its streamlined nature, is processed virtually in real-time.

Even after achieving acceptable levels, delays in packet networks are still higher than in circuit-switched networks and require echo cancellation regardless of distance.

The characteristics of signaling in packet networks are explored in the next section.

What Drives VoIP. The trend toward the use of packet networks for voice communication is driven mostly by inherent technical and economical factors, although regulatory factors (such as lower access charges) can also play a role. An obvious advantage is the cost savings achievable with one converged network compared to maintaining separate voice and data networks, especially since bandwidth demand for data is growing much faster than for voice. Another advantage is the ability to introduce next-generation services that blend voice and data in a simple and cost-effective way. The final advantage is bandwidth efficiency. It may sound counterintuitive, at first, that bandwidth efficiency is achievable with VoIP, since voice samples are burdened with several layers of headers (and trailers) in order to be transported. But bandwidth can be saved, in spite of the added overhead, with two techniques:

1. Voice compression
2. Silence suppression

Voice sources typically sample the analog signal at 8 kHz and, in the SCN, encode it into 8-bit samples that are sent at 64 kb/s. The SCN is built around the 64-kb/s rate and cannot easily take advantage of voice compression, which now allows good quality voice at rates from 8 to 13 kb/s. In a packet network, compression provides immediate bandwidth reduction because a group of (sequential) voice samples are shipped in a packet only when the packet becomes available, regardless of the bit rate of the data link. Furthermore, since the payload is passed unmodified through the network, voice coding affects only source and destination, which can negotiate coding standards via signaling messages. That avoids format conversion inside the network (a source of delay) and facilitates the introduction of new compression techniques.

Silence suppression (or *discontinuous transmission*—DTX) uses *voice activity detection* (VAD) techniques, which can achieve bandwidth savings of around 40% by not sending any packet during silent pauses.

20.3.2 VoIP Networks

To understand how voice connections operate in VoIP networks, let us consider the two basic components of the switched-circuit network (SCN)—

- Communication channels
- Telephone exchange

and discuss how their functions translate into components of a packet network.

Communication Channels. In the SCN, communication channels are physical entities, such as electrical circuits, frequency bands, or time slots. In a packet network, channels are logical entities based essentially on the L3 and L4 addresses of the source and the destination. In an IP network, those addresses are the IP address and the TCP/UDP port (the socket). Two peer applications establish a logical channel between them when they agree to use a specific pair of sockets (one for the origination, one for the destination) for the duration of the information flow. A logical channel is an application layer (L7) construct that uses the services and works "on top" of the L4 connection (if any). It can use either TCP or UDP for transport and is established by an exchange of signaling messages that carry the parameters of the channel. Logical channels can be unidirectional or bidirectional and can be of two types, based on the information they carry (Section 20.3.1):

1. Media, or bearer, channels
2. Signaling channels

In the SCN the two channel types form two separate networks: the voice-trunk network and the SS7 signaling network. A packet network carries both types of channels, indistinguishable except for the "sockets" in the packet headers.

Telephone Exchange. In a VoIP network, the SCN component known as the telephone exchange disappears as a monolithic, integrated unit and is decomposed into functional elements that can be implemented as separate physical units interconnected by the packet network (Fig. 20.3-1). The result is a distributed, modular architecture where functional elements communicate via signaling messages and their physical location is, at least in principle, irrelevant.

Two architectural approaches have emerged:

1. Centralized control
2. Distributed control

Centralized control is called the *server-routed model* in ITU-T literature and has evolved from SCN architecture. In this approach call control is centralized in servers containing call logic and terminals have no call processing intelligence. Terminals do not exchange signaling messages directly with one another but only with call servers, using *master–slave protocols*.

Distributed control is called the *direct routed model* in ITU-T literature and has evolved from Internet architecture, extending it to include voice communication. In this approach packet endpoints, such as VoIP terminals, contain call logic and exchange signaling messages with each other using *peer-to-peer* protocols. Servers play a supporting role for routing and access control and may help with call control for complex services.

Typical VoIP implementations use a mixed approach: centralized control for interfaces to the SCN and distributed control for packet terminals. As a result,

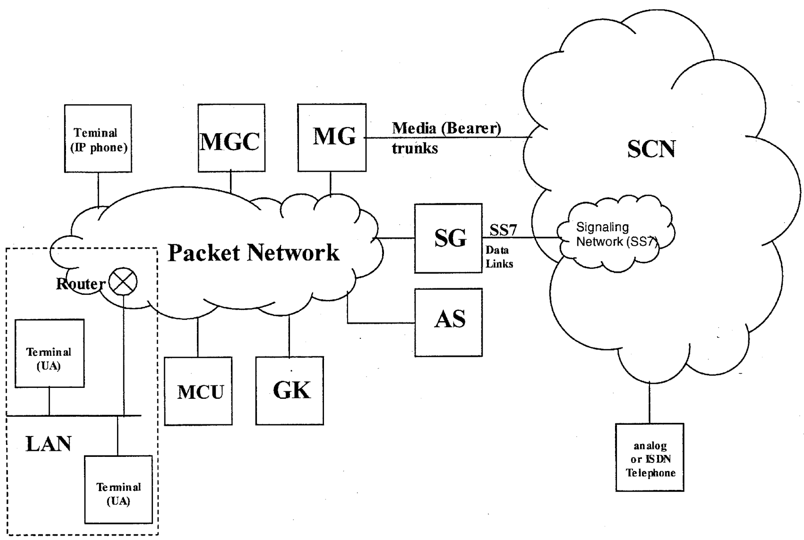


Figure 20.3-1 Functionally decomposed VoIP elements in a packet network.

some VoIP components support the centralized approach, some support the distributed approach, and some support both, as will be explained in more detail at the end of this section.

Regardless of the approach, the role of the *switchblock* (or *switch fabric*—Section 1.7.2), traditionally an integrated part of the telephone exchange, is fulfilled by the packet network, which provides a global fabric capable of connecting user terminations and VoIP functional elements.

Functional Elements of a VoIP Network. VoIP network elements, described below, are all *packet endpoints* because they originate or terminate the packet portion of a connection.

Gatekeeper (GK). This element is also called *proxy server*. GKs (proxies) provide access control and address translation, acting as checkpoints for access to the network and as routing intermediaries between originating and terminating end points. They may use the services of application servers (see below). A GK manages a set of packet terminals (called a *zone* or a *domain*) and is a member of an *administrative domain*, composed of all the zones or domains under single administrative control. That leads to a hierarchical structure where some GKs provide interfacing services between administrative domains to lower ranking GKs.

Media Gateway (MG). This element is sometimes called *voice gateway* or simply *gateway*. A MG provides the interface between dissimilar formats or standards for media streams, under the control of a media gateway controller (MGC—see below). A MG typically connects lines and/or trunks in the SCN to bearer (media) channels in the packet network.

The term voice gateway is used when a MG handles only voice communication. Other terms, such as *access gateway, residential gateway*, or *trunking gateway*, are often used with a loosely defined meaning that refers to the type of SCN circuits served by the MG.

A media gateway that connects to an access network as defined in Section 1.8, thus interfacing with subscriber lines, digital concentrators, and access systems, is called an access gateway or a residential gateway. Such a gateway has the ability to detect signaling events at the access interfaces and to report them to the MGC; the MGC, in turn, instructs the MG to deploy signals at those interfaces. The term access gateway is occasionally used also as a general term for media gateway, with reference to the fact that it provides access to the packet network from the SCN and vice versa.

A media gateway that connects to SCN trunks is called a trunking gateway. If the trunks use common-channel signaling, the trunking gateway connects only to voice trunks, and signaling data links are connected to a signaling gateway (see below). For channel-associated (CAS) trunks, the trunking gateway must detect and apply signaling from/to the trunks, under control of the MGC.

Signaling Gateway (SG). This element is sometimes called *telephony gateway*. The SG manages the interface between the SS7 signaling network of the SCN and the packet network. It terminates signaling data links to/from the SCN and behaves like a STP or a SEP (Chapter 5). A SG is controlled by a media gateway controller (see next paragraph). This element is most often combined with a GK or a MG in implementations.

Media Gateway Controller (MGC). This element is sometimes called *call agent* or *softswitch*. A MGC controls one or more MGs and SGs and contains the equivalent of the call processing logic of a telephone exchange. The term softswitch is somewhat ambiguous, since it is often used as a marketing term for units that combine a MGC with a signaling gateway and/or an application server (see below).

User Agent (UA). This element is similar in function to a MGC, but it controls (and usually resides in) a single user endpoint such as a packet terminal. It contains call processing logic and acts on behalf of a user, like a calling and/or a called party. In this and in the next chapter, the term "packet terminal" is often used to mean the "user agent controlling the packet terminal."

Multipoint Control Unit (MCU). This element supports multiparty conferences.

Application Server (AS). This element provides the functions that, in the SCN, reside in a SCP or in a NMS, that is, intelligent network service logic and network management logic. An AS may also host, under control of a GK, applications known collectively as AAA (authentication, authorization, and accounting).

MGCs, MGs, and SGs are used in centralized architectures, that is, primarily in support of SCN interfaces; GKs and UAs are used in distributed architectures, that is, primarily in support of packet terminals; MCUs and ASs are used in both architectures. It should be noted that, in principle, either architecture could be used to support both packet terminals and SCN interfaces. For instance, packet terminals could be envisioned as single-port MGs (with no call logic) controlled by a centralized MGC. Conversely, each SCN interface (line or trunk) could be controlled by its own UA. But the economics are not in favor of such nontypical approaches.

GKs, MGs, MGCs, SGs, and ASs may be implemented as separate physical units but commonly are grouped in different combinations, such as a MG/SG unit, or a softswitch with MGC, SG, and AS functions.

20.3.3 Characteristics of Signaling Protocols in VoIP Networks

Signaling protocols in VoIP networks have more granularity compared to SCN signaling, an approach called "one function, one protocol." Each layer of the OSI reference model (and each sublayer, when present) usually corresponds to a separate protocol. Bearer channels and signaling channels use different protocols and different protocols may be used to perform different tasks during the same call. The term "protocol suite" is used to denote a set of VoIP protocols that work in concert.

On the other hand, the plethora of country-specific protocol variations (typical in the SCN) has not materialized to any appreciable extent, due to the global nature of VoIP.

In a VoIP network, all the functional elements that result from the decomposition of the telephony exchange must be viewed as hosts (computers) communicating with each other via protocols structured according to the OSI reference model. Even analog or ISDN phones can be viewed as hosts since they are connected to the packet network via hosts (MGs and SGs). As a host, each functional element operates at all layers of the OSI model, up to the application layer (L7) and, in this chapter, we will assume that L5 and L6 functions are included in L7. Signaling is a function of the application layer. For the lower layers, VoIP uses the TCP/IP protocol suite at L3 and L4 and TCP/IP-compatible protocols at L1 and L2. For reliability reasons, most signaling applications use TCP (rather than UDP) as the L4 protocol.

How VoIP Works. Calls in a VoIP network are set up by means of signaling messages exchanged on logical channels between VoIP functional elements, logical channels that are themselves established by signaling messages. For the purpose of understanding VoIP operation, let us consider a conceptual call scenario where the originating user (A) is directly connected to the packet network, and the terminating user (B) is connected to the SCN (Fig. 20.3-2). User A invokes the VoIP application in its user agent (UA-A) and provides it with the address of B. If B were directly connected to the packet network, A could input B's 32-bit IP address. Since B is connected to the SCN, B is identified by a telephone number, so we will assume that the application in UA-A accepts telephone numbers. UA-A establishes a signaling channel with GK-A, which has a known address, and sends B's number to it. With the help of GK-A's routing logic (which translates B's telephone number into the

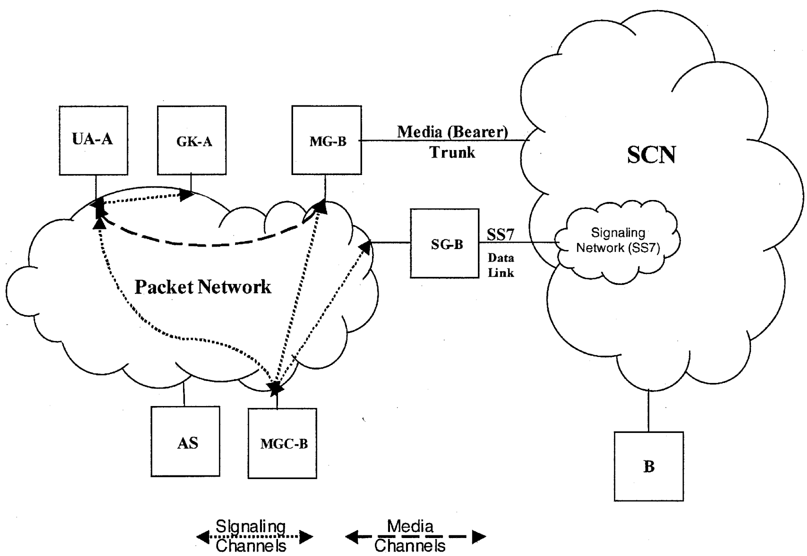


Figure 20.3-2 Conceptual example of a VoIP call.

IP address needed for the call), UA-A establishes a signaling channel with MGC-B, after which the channel to GK-A may be released. After receiving destination information from UA-A, MGC-B establishes signaling channels with SG-B and MG-B to get access to the SCN. Through SG-B and MG-B, MGC-B sets up a circuit-switched voice connection from MG-B to B and then relays alerting information, received from B via SG-B, to UA-A. UA-A and MGC-B set up two unidirectional media channels (one per direction) between UA-A and MG-B, which MG-B connects to the SCN voice trunk going to B. When B answers, the answer is relayed from B to SG-B, to MGC-B, to UA-A. The media connection is then activated and voice streaming starts. At release, UA-A informs MGC-B of the release (or vice versa depending on which party hangs up first) with a message on the signaling channel, then both release the media channels and the signaling channels.

Examples of call sequences involving specific signaling protocols are provided in the next chapter.

20.4 LOWER LAYER PROTOCOLS FOR VoIP

We have seen that VoIP uses UDP for media transport (because of delay constraints) and mostly TCP for signaling transport (because of reliability). Two additional L4 protocols have been designed specifically for VoIP: one for media streams (RTP) and the other for signaling (SCTP). They address some drawbacks of UDP and TCP. Another protocol, SDP, is used to negotiate media (bearer) characteristics.

20.4.1 The Real-Time Transport Protocol (RTP)

UDP, while meeting the timing requirements for real-time traffic, is not sufficient for good quality media-stream connections because it cannot ensure ordered delivery of packets. The most commonly used protocol to fill that gap is RTP [21]. RTP's main function is sequence numbering, which is used by the application at the receiving end to reconstruct the packet sequence, to detect packet loss, or to determine the location of a packet in video communication. RTP is independent of the underlying transport protocol but it typically sits "on top" of UDP in the protocol stack and adds an extra header to UDP datagrams. The RTP standard contains a degree of flexibility that allows it to be customized, as documented in ancillary documents called *profiles*. Profiles are also used for audio/video encoding formats and for protocol extensions. RTP PDUs are called *packets*.

The protocol has two components: RTP proper, which carries the media stream, and RTP Control Protocol (RTCP), used by session participants to exchange packets for control and monitoring of the media streams. More specifically, RTCP provides the following:

- Feedback on the quality of data distribution to all the parties involved in a session (two-party or multiparty call)
- A stable identifier for a RTP source called *canonical name* (CNAME)

- Control of the rate at which each participant sends RTCP packets to all others based on the number of participants in the session, to avoid overloading the network

Any two session participants that exchange RTP media streams also exchange RTCP packets. Multiple RTCP packets are carried in one UDP (or equivalent L4 protocol) segment; the minimum number is two per segment.

RTP and RTCP streams in the same session use two sequential UDP ports: an even number for RTP and the next higher (odd) number for RTCP.

RTP Packet Format. The fields in the RTP header (Fig. 20.3-3) are as follows:

- *M* (*Marker*): Used optionally for PDU delineation when more than one RTP PDU is carried in a UDP segment.
- *CC* (*CSRC Count*): The number of CSRC identifiers (see below).
- *PT* (*Payload Type*): Payload format used by the receiver to interpret the payload.
- *Sequence Number*: Starts as a random number and is incremented by one for each packet sent.
- *Timestamp*: Sampling time of the first octet of the payload, based on a linear clock. It is used by the destination to calculate jitter and synchronize with the source.
- *SSRC* (*Synchronization Source*) *Identifier*: Identifies the synchronization source.
- *CSRC* (*Contributing Source*) *Identifiers*: A list of SSRC identifiers for contributing sources, such as multiple parties in a conference call that are contributing to the packet.

The header is followed by the payload, whose format depends on the application. Payload formats are documented in profiles and are not covered in this book (nor in the RFC).

RTCP Packet Format. RTCP packets have formats that vary depending on the payload. The header is similar to the RTP header (Fig. 20.3-4) and contains either

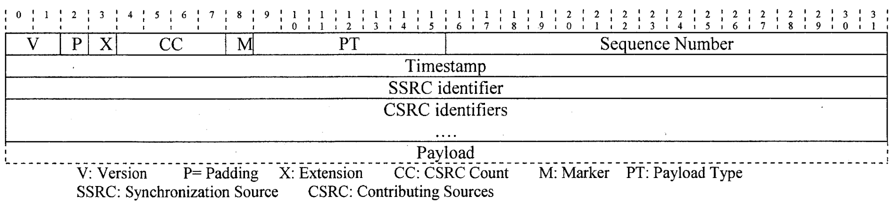


Figure 20.3-3 RTP header format. (Source: RFC 3550 [16].)

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
V		P	(*)					PT								Length															
SSRC/CSRC																															
Payload																															

V: Version P: Padding *: depends upon the type of packet PT: packet type
SSRC: Synchronization Source CSRC: Contributing Source

Figure 20.3-4 RTCP header format. (Source: RFC 3550 [16].)

SSRC or CSRC based on the type of packet. The header is followed by variable-length elements, all ending at the 32-bit boundary, whose formats depend on the type of information they carry. One type of packet (*source description*—SDES) carries, in the variable-length part, the CNAME parameter needed by receivers because SSRC may change during a session. CNAME is used also to synchronize multiple streams (e.g., voice and video) from the same source. More details can be found in [21].

20.4.2 The Stream Control Transmission Protocol (SCTP)

SCTP is a transport layer (L4) protocol in the same class as TCP and UDP, with which it shares several aspects. SCTP is designed as an alternative to TCP and UDP for transporting signaling messages, although nothing prevents it from being used for other types of applications [22,23]. TCP or UDP can be (and are) used for transporting signaling messages, but they are not optimized for that role. TCP provides reliability and sequencing but is bit-stream-oriented rather than message-oriented; thus it does not support message delineation. UDP is message-oriented but does not provide reliability or sequencing. SCTP provides the reliability and sequencing of TCP with the message orientation of UDP. It also addresses other TCP drawbacks, such as:

- *Head of line* delays, which occur when the loss of a segment stops delivery of all segments behind it (in the same connection) during retransmission
- Vulnerability to network-level faults
- Vulnerability to attacks such as *denial of service*

SCTP is connection-oriented through the use of *associations*, similar to TCP connections. An association is established between two SCTP entities before user data is exchanged and is shut down at the end of the data transfer. It guarantees reliable delivery and sequencing of packets, as well as flow and congestion control, with a windowing and retransmission mechanism like TCP's.

TCP drawbacks are addressed by enhanced functionality, such as:

- Message-based transmission with *bundling* and *fragmentation*
- Multistreaming
- Multihoming
- Option for unsequenced delivery, when only reliable delivery is needed

SCTP uses the term *stream* for a unidirectional sequence of messages, and a SCTP association can contain multiple streams. Multistreaming eliminates the head-of-line bottleneck since even when a stream is halted during retransmission the other streams in the association continue sending. An example of multistreaming is when the messages of a SS7 signaling data link (SDL) connected to a SG are transported to a processing point in the packet network: the message flow for the entire SDL is transported as one association but the messages related to each trunk served by the SDL are assigned to their own stream. With that arrangement a delay in one stream will only affect one call.

Another feature of SCTP is the support of *multihoming* associations. Multihoming means that the source or the destination (or both) of an association can be identified by multiple IP addresses. One of the addresses, the primary address, is always known in advance and is used for starting the setup; the others may be exchanged during setup. If one address is unreachable or becomes unreachable, another address is used to establish or continue communication. This functionality can provide reliability against network failures if, for example, each address is assigned to a separate port in a node and is set up so that a connection using it will follow a physically different path from the others.

Protection against security attacks is addressed in two ways:

- A *verification tag* in the packet header
- A *cookie exchange* during association setup

A verification tag's value is chosen by each of two communicating nodes during startup, and packets not carrying the agreed-to value are discarded. The cookie exchange is discussed later in this section.

SCTP PDUs are called *packets* and their format is shown in Fig. 20.4-1. Like TCP and UDP, SCTP uses port numbers to identify the transported application. The payload of a SCTP packet is carried in a series of *chunks* (Fig. 20.4-2), containing control or user data and providing message delineation. Control chunks are used to manage the SCTP association; user-data chunks carry user messages (e.g., a signaling message—the SCTP user is a signaling application). Multiple chunks can be

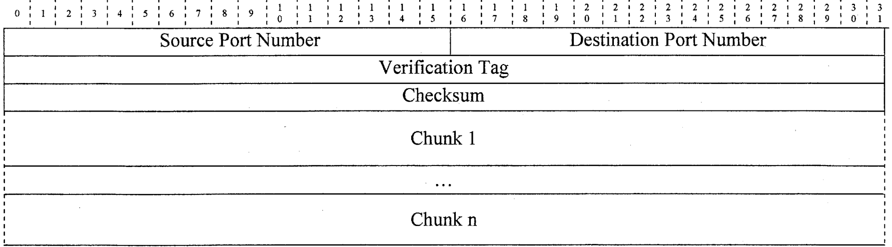


Figure 20.4-1 SCTP packet format. (Source: RFC 2960 [18].)

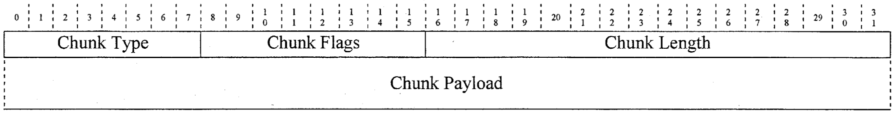


Figure 20.4-2 SCTP chunk format. (Source: RFC 2960 [18].)

bundled into a packet, but if a message is longer than the MTU of the path, it can be fragmented into multiple packets. Some control chunks cannot be bundled and for the details we refer the reader to [23]. There are 13 currently defined chunk types (Table 20.4-1); one (DATA) is for user payloads, and the others are for control. The format of the DATA chunk is shown in Fig. 20.4-3. Reliable delivery and sequencing are handled separately: every chunk, even when out of sequence, is acknowledged based on the value of TSN, while SSN is used to sequence chunks within a stream. Acknowledgments and retransmission are managed via control chunks. The option of providing reliability with no sequencing is controlled by the U (Unsequenced) bit in the chunk header. Bits B (Beginning) and E (End) are used for fragmentation.

Associations are established with a *four-way handshake* (Fig. 20.4-4), similar to the three-way handshake of TCP, but with the addition of the cookie exchange. The cookie is a data package that is generated by the recipient of the INIT request and is included in the INIT ACK chunk.

TABLE 20.4-1 Chunk Types

Chunk Type	Acronym	Code
Payload data	DATA	0
Initiation	INIT	1
Initiation acknowledgment	INIT ACK	2
Selective acknowledgment	SACK	3
Shutdown association	SHUTDOWN	7
Shutdown acknowledgment	SHUTDOWN ACK	8
Cookie echo	COOKIE ECHO	10
Cookie acknowledgment	COOKIE ACK	11
Shutdown complete	SHUTDOWN COMPLETE	14

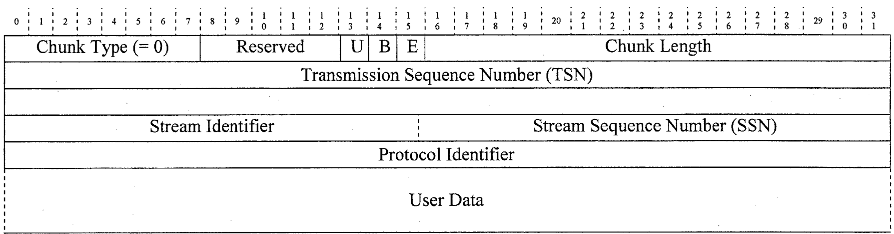


Figure 20.4-3 DATA chunk format. (Source: RFC 2960 [18].)

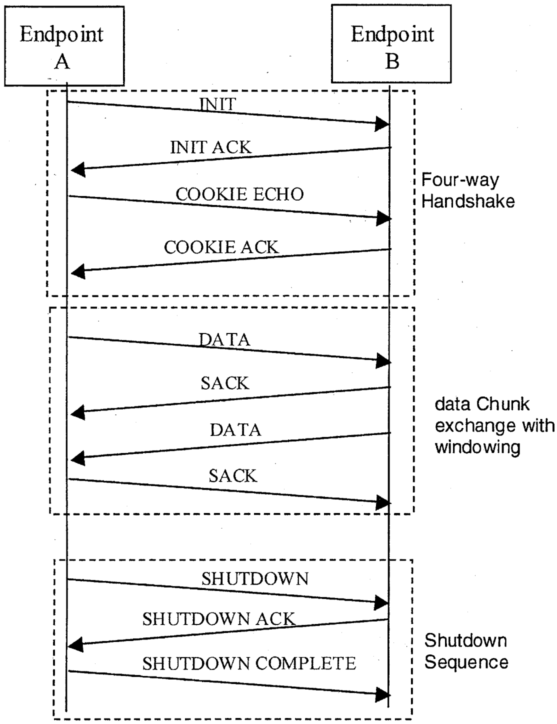


Figure 20.4-4 Association setup and shutdown sequence.

An association is shut down with the shutdown procedure. Congestion and flow control work the same way as for TCP, with the window value (in bytes) initially established by the sender in the INIT chunk. Windows apply to associations and not to individual streams.

SCTP Users. An application using SCTP is called an *upper layer protocol* (ULP). ULP is a generic term, which reflects the fact that SCTP can be used as a general-purpose transport layer protocol.

20.4.3 The Session Description Protocol (SDP)

The SDP protocol [24] is used to carry descriptions of multimedia sessions, which may involve voice, video, and whiteboard streams. The protocol covers only descriptions and not the negotiation of the session characteristics, which is the responsibility of the signaling application. SDP must work in concert with the signaling protocols used for session setup, such as SIP (Chapter 21).

SDP is a text-based protocol: a session description consists of a series of lines of text, each of which has the basic structure

$$type = value$$

where *type* is a one character token and *value* is one or more parameters separated by spaces. A session description has two parts, each consisting of multiple lines: a *session-level part* and a *media-level part*.

Session-level part. This part starts with a "v=" line and contains information that applies to the whole session, unless overridden by a line in the media-level part (below). This part consists of a few mandatory lines followed by optional lines.

Media-level part. This part starts with an "m=" line. This part is optional and, if present, contains descriptions of one or more individual media streams in the same session, each starting with an "m=" line.

The most significant lines in a session description are shown in Table 20.4-2. The order in which lines appear in a session description is fixed and must be as it appears in the table. The parameter *fmt_list* in the "m=" line is the list of formats for the media (such as the media payload types defined in RTP profiles) that can be used for the session. The first item in the list is the default.

TABLE 20.4-2 SDP Session Description Format

	Content	Format[a]		Note
Session-level part	SDP version	**v** = *version_number*	M	
	Origination data	**o** = *username session_id version network_type address_type address*	M	
	Session name	**s** = *session_name*	M	
	Session information	**i** = *session_description*	O	1
	URI (for additional session information)	**u** = *URI*	O	
	Connection data	**c** = *network_type address_type connection_address*	O	2
	Start and stop time	**t** = *start_time stop_time*	M	3
	Attributes	**a** = attribute or **a** = attribute:value	O	4
Media-level part	Media stream	**m** = *media port transport fmt_list*	M	5
	Connection data	**c** = *network_type address_type connection_address*	O	2
	Attributes	**a** = *attribute* or **a** = *attribute: value*	O	4

[a]Items in **bold** are as they appear in the session description; items in *italics* are names of parameters.
M = mandatory and O = optional.

Notes:

1. A free format description of the session.
2. This is actually a mandatory field, in the sense that it must appear either in the session-level part or in the media-level part.
3. One or more of these lines are allowed if the session is active at different but nonregular times.
4. Attributes are used mainly to extend SDP, that is, to add new functionality with backward compatibility without issuing a new version of the specification. Multiple attribute lines are allowed.
5. This line is mandatory only within a media-level stream description (which is optional).

20.5 ACRONYMS

AAA	Authentication, authorization, and accounting
AS	Application server
ATM	Asynchronous transfer mode
CC	CSRC count
CNAME	Canonical name
CSRC	Contributing source
DNS	Domain name system
DoD	Department of Defense
ES	End system
FQDN	Fully qualified domain name
GK	Gatekeeper
GW	Gateway
H	Header
HDLC	High-Level Data Link Control
HTTP	Hypertext Transfer Protocol
ID	Identity
IETF	Internet Engineering Task Force
IP	Internet Protocol
IS	Intermediate system
ISDN	Integrated Services Digital Network
ISO	International Standards Organization
ISUP	ISDN User Part
ITU	International Telecommunication Union
ITU-T	Telecommunication Standardization Sector of ITU
L2	Layer 2
L3	Layer 3
L4	Layer 4
L5	Layer 5
L6	Layer 6
L7	Layer 7
LAN	Local area network
LLC	Logical link control
M	Marker
MAC	Medium access control
MCU	Multipoint control unit
MG	Media gateway
MGC	Media gateway controller
MTP	Message Transfer Part

MTU	Maximum transmission unit
NGN	Next-generation network
NMS	Network management system
OSI	Open Systems Interconnection
PC	Point code
PDU	Protocol data unit
PLC	Packet loss concealment
PPP	Point-to-Point Protocol
PSTN	Public switched telephone network
PT	Payload type
QoS	Quality of service
RFC	Request for comments
RTCP	RTP Control Protocol
RTP	Real-Time Transport Protocol
SCCP	Signaling Connection Control Part
SCN	Switched-circuit network
SCP	Service control point
SCTP	Stream Control Transmission Protocol
SDP	Session Description Protocol
SDU	Service data unit
SEP	Signaling end point
SG	Signaling gateway
SIP	Session Initiation Protocol
SS7	Signaling System No. 7
SSRC	Synchronization source
STP	Signal transfer point
T	Trailer
TCP	Transmission Control Protocol
TDM	Time-division multiplexing
TSAP	Transport service access point
UA	User agent
UDP	Used Datagram Protocol
URI	Uniform resource identifier
URL	Uniform resource locator
VoIP	Voice over IP
WAN	Wide area network
WWW	World Wide Web

20.6 REFERENCES

1. D. E. Comer, *Internetworking with TCP/IP Volume 1.Principles, Protocols and Architectures*, Prentice Hall, Upper Saddle River, NJ, 2000.
2. W. Stallings, *Data and Computer Communications*, Prentice Hall, Upper Saddle River, NJ, 2003.
3. C. E. Spurgeon, *Ethernet: The Definitive Guide*, O'Reilly & Associates, Sebastopol, CA, 2000.
4. *Hypertext Transfer Protocol—HTTP/1.1*, RFC 2068, IETF, Jan. 1997.
5. *The Point-to-Point Protocol (PPP*), RFC 1661, IETF, July 1994.
6. *Internet Protocol DARPA Internet Program Protocol Specification*, RFC 791, IETF, Sept. 1981.
7. *Internet Protocol Version 6 (IPv6) Specification*, RFC 2460, IETF, Dec. 1998.
8. *Internet Standard Subnetting Procedure*, RFC 950, IETF, Aug. 1985.
9. *Classless Inter-Domain Routing (CIDR): An Address Assignment and Aggregation Strategy*, RFC 1519, IETF, Sept. 1993.
10. *Traditional IP Network Address Translator (Traditional NAT)*, RFC 3022, IETF, Jan. 2001.
11. *Domain Names—Concepts and Facilities*, RFC 1034, IETF, Nov. 1987.
12. *Domain Names—Implementation and Specification*, RFC 1035, IETF, Nov. 1987.
13. *Transmission Control Protocol DARPA Internet Program Protocol Specification*, RFC 793, IETF, Sept. 1981.
14. *User Datagram Protocol*, RFC 768, IETF, Aug. 1980.
15. *Uniform Resource Identifiers (URI): Generic Syntax*, RFC 2396, IETF, Aug. 1998.
16. *Uniform Resource Locators (URL)*, RFC 1738, IETF, Dec. 1994.
17. J. Ellis, C. Purcell, and J. Rahman, *Voice, Video, and Data Network Convergence*, Academic Press, San Diego, CA, 2003.
18. R. Swale et al., *Voice Over IP Systems and Solutions*, The Institution of Electrical Engineers, London, 2001.
19. J. F. Durkin, *Voice-Enabling the Data Network: H.323, MGCP, SIP, QoS, SLAs, and Security*, Cisco Press, Indianapolis, IN, 2003.
20. U. Black, *Voice Over IP*, Prentice Hall PTR, Upper Saddle River, NJ, 2002.
21. *RTP: A Transport Protocol for Real-Time Applications*, RFC 3550, IETF, July 2003. Copyright (c) The Internet Society (2003). All rights reserved. This document and translations of it may be copied and furnished to others, and derivative works that comment on or otherwise explain it or assist in its implementation may be prepared, copied, published and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice and this paragraph are included on all such copies and derivative works. However, this document itself may not be modified in any way, such as by removing the copyright notice or references to the Internet Society or other Internet organizations, except as needed for the purpose of developing Internet standards in which case the procedures for copyrights defined in the Internet Standards process must be followed, or as required to translate it into languages other than English. The limited permissions granted above are perpetual and will not be revoked by the Internet Society or its successors or assigns.

22. *An Introduction to the Stream Control Transmission Protocol (SCTP)*, RFC 3286, IETF, May 2002.
23. *Stream Control Transmission Protocol*, RFC 2960, IETF, Oct. 2000. Copyright (c) The Internet Society (2000). All rights reserved. (The rest of the copyright statement is the same as in [21].)
24. *SDP: Session Description Protocol*, RFC 2327, IETF, Apr. 1998.

21

SIGNALING FOR VoIP

21.1 INTRODUCTION

The specifications for VoIP signaling protocols consist of a large set of documents, totaling thousands of pages. This chapter condenses those specifications into some thirty pages and is therefore limited to a relatively high-level look at the most important signaling protocols, namely:

- H.323
- SIP
- Gateway Control Protocol/H.248
- SIGTRAN
- BICC

Each of these protocols performs a specific function and can be used for signaling between certain functional elements (Section 20.3.2). The most typical combinations are shown in Table 21.1-1 and are discussed in more detail in the introductory section of each protocol. H.323, SIP, and BICC are the primary call-setup protocols; each of the other protocols performs auxiliary functions and works in concert with one of the main three. In this chapter a VoIP network that uses either H.323, SIP, or BICC as the primary call-setup signaling protocol is referred to as a H.323, SIP, or BICC network, respectively.

Figure 21.1-1 shows a deployment scenario of VoIP signaling protocols for connections where the origination and/or the termination is a packet terminal (UA), and the primary call-setup protocol is either H.323 or SIP. Figure 21.1-2 shows a deployment scenario where both the origination and the termination are in the switched-circuit network (SCN) and the primary call-setup protocol is BICC. Each scenario reflects one of the two avenues of convergence for voice and data discussed at the beginning of Section 20.3.

Signaling in Telecommunication Networks, Second Edition, by John G. van Bosse and Fabrizio U. Devetak

TABLE 21.1-1 VoIP Signaling Protocols Versus Network Elements

	GK	MG	MGC	UA	SG
GK	H.323, SIP	NA	H.323, SIP	H.323, SIP	NA
MG	NA	RTP[a]	H.248	RTP[a]	NA
MGC	H.323, SIP	H.248	H.323, SIP, BICC	H.323, SIP	SIGTRAN
UA	H.323, SIP	RTP[a]	H.323, SIP	H.323, SIP, RTP[a]	NA
SG	NA	NA	SIGTRAN	NA	SIGTRAN

[a] RTP (Section 21.4.1) is not a signaling (L7) protocol and is used to transport packets in media streams, not for call control.

Note: NA, not applicable.

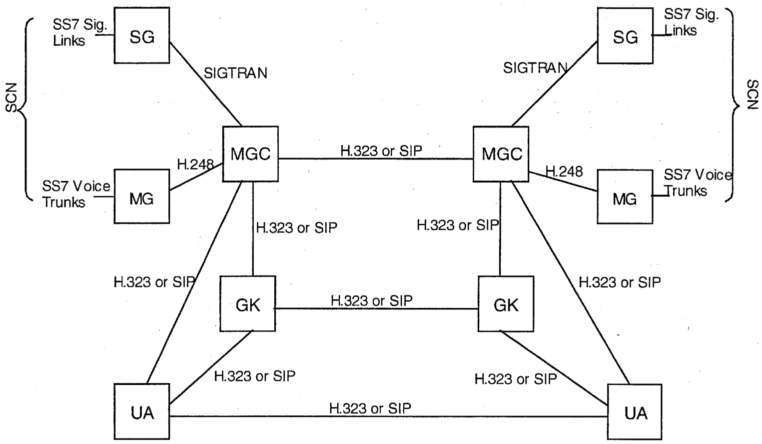


Figure 21.1-1. VoIP signaling protocols and their relationships in H.323 or SIP networks.

VoIP signaling protocols are application layer (L7) protocols that manage call control. They utilize the services of TCP, SCTP, UDP, and SDP at Layer 4, of IP at Layer 3, and of protocols like Ethernet or PPP at Layer 2.

Media streams typically use the services of UDP and RTP at Layer 4, of IP at Layer 3, and Ethernet or PPP at Layer 2. Their Layer 5–7 functions are handled by applications that are not discussed in this book.

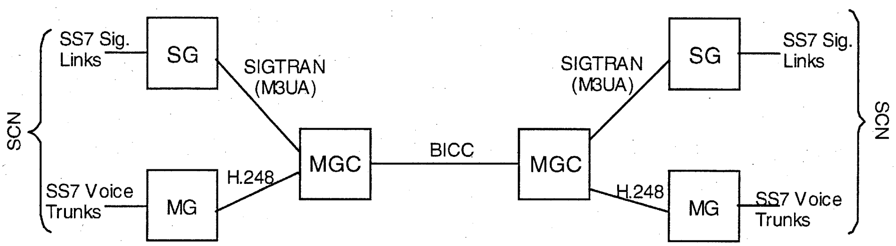


Figure 21.1-2. VoIP signaling protocols and their relationships in BICC networks.

21.2 THE H.323 PROTOCOL

21.2.1 Overview

H.323 is an ITU-T recommendation that specifies standards for VoIP [1]. The recommendation was first introduced in 1996 and initially covered requirements for multimedia conferencing of packet terminals connected to LANs that offer no guaranteed quality of service (QoS). The recommendation was later extended to cover voice communication over IP networks that may not provide guaranteed QoS. This chapter follows Version 4 of the recommendation.

H.323 has its conceptual roots in classic telephony and it has developed a reputation for being somewhat complex and inflexible. In spite of that, it has enjoyed widespread deployment. H.323 is a complete standard for multimedia communication in IP networks and is called an *umbrella specification* since it links together and incorporates a large number of other specifications, not all issued by ITU-T.

H.323 consists of a suite of protocols based on the direct-routed model (distributed control) and is generally used for peer-to-peer signaling between UAs and between UAs and MGCs. Figure 21.2-1 shows an example of a network that uses H.323 components and signaling. Figure 21.2-2 shows the protocol stack.

21.2.2 H.323 Structure

The network components defined in the H.323 recommendations (Figs. 21.2-1 and 21.2-3) are fairly consistent with the VoIP network elements introduced in Chapter 20:

- Terminal (UA)
- Gatekeeper (GK)
- Gateway (GW)

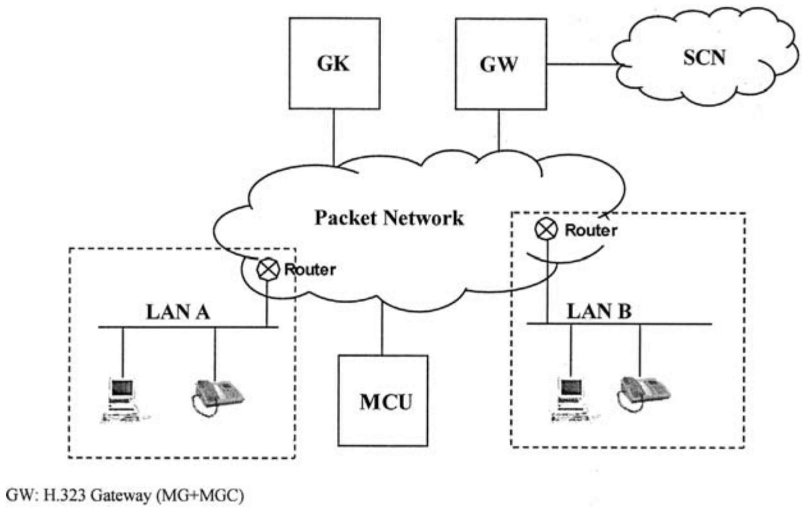


Figure 21.2-1. Example of a H.323 network.

Video Codecs	Audio Codecs	RAS Channel	Call Signaling Channel	Control Channel	Data Communication
RTCP & RTP		H.225 - RAS	H.225 – Q.931	H.245	T.120
UDP			TCP		
IP					
Data link					
Physical					

Figure 21.2-2. H.323 protocol stack.

- MCU/MC
- Border element (BE)

Terminals and gateways are H.323 *endpoints*, because they originate and/or terminate a VoIP connection. Two types of terminals are supported: full-featured terminals, and *simple endpoint type* (SET) terminals specified in Annex F of Rec. H.323. SET terminals are simple, single-purpose terminals with limited functionality, such as voice-only packet telephones.

Gatekeepers are recommended, but not mandatory, components of an H.323 network. In addition to performing standard functions such as access control and routing, they provide bandwidth control and, optionally, some form of call

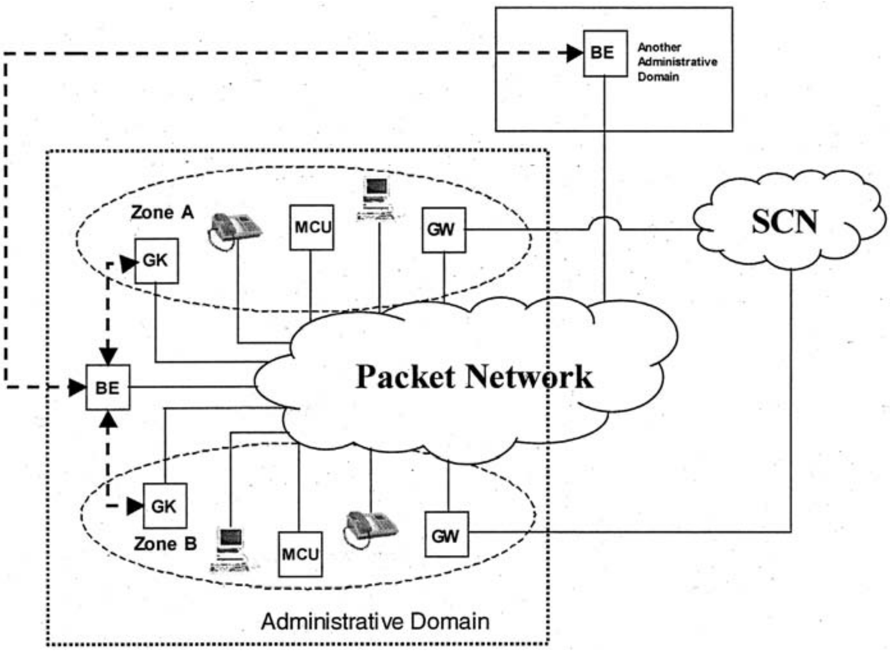


Figure 21.2-3. H.323 Network spanning multiple administrative domains.

TABLE 21.2-1 H.323 Recommendation Summary

Sponsor	Standard	Signaling Function
ITU-T	H.225.0	Endpoint registering with gatekeeper (RAS protocol) Call setup and tear-down (Q.931 protocol)
ITU-T	H.245	Channel parameter negotiation
IETF	RFC 1889, 1890	Media stream transport (RTP)
IETF	RFC 1889, 1890	Media stream control (RTCP)
ITU-T	H.235	Security
ITU-T	H.332	Conferencing
ITU-T	H.450.1, H.450.2, H.450.3	Supplementary services
ITU-T	G.711, G.722, G.721.1, G.723.1, G.728, G.729,	Audio codec standards
ITU-T	H.261, H.263	Video codec standards
ITU-T	T.38	Fax service
ITU-T	T.120	Data service
ITU-T	T.122, T.124, T.126, T.127	Data communications

control. The set of terminals, gateways, and MCUs managed by one GK is called a *zone*. Multiple zones may be grouped into an *administrative domain*.

The border element (BE) is a type of gatekeeper that works as a collection point for the GKs in an administrative domain and interfaces with BEs in other administrative domains to provide routing and security for interdomain communication.

H.323 gateways are defined as multifunctional elements that combine MG, MGC, and SG functions.

The standards that are covered by Rec. H.323 are summarized in Table 21.2-1. The core functions are provided by the following protocols:

- Registration, admission, and status (RAS)
- Q.931
- H.245
- RTP/RTCP

H.323 Channels. An H.323 call involves messages transported in multiple independent channels called *information streams*, each of which uses one the protocols listed above (Fig. 21.2-4).

RAS Channel. This channel transports RAS protocol messages (Rec. H.225 [2]) and provides signaling communication between an H.323 endpoint and a GK. Endpoints use this channel to discover a GK, to register with it, and to gain admission to the packet network for a call. A GK uses this channel to control the bandwidth used for a call, to open a Call Signaling Channel, and to control access to the network. This channel uses UDP transport.

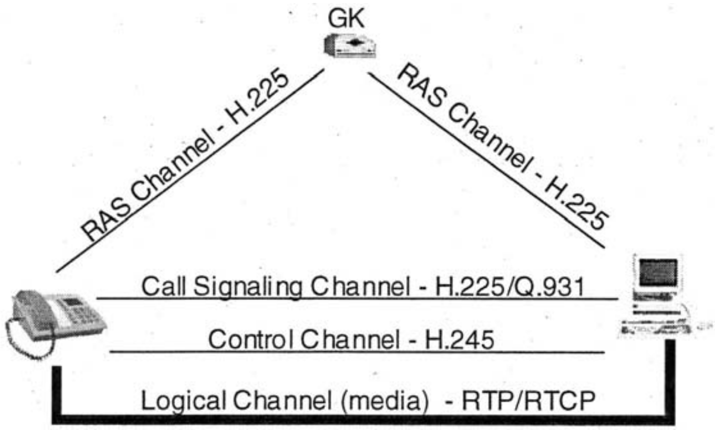


Figure 21.2-4. H.323 channels.

Call Signaling Channel. This channel transports Q.931 protocol messages (Chapter 10), as amended by Rec. H.225. It provides signaling communication between H.323 endpoints for basic call-control messages (call setup and release) and to obtain the transport address of the control channel. It uses TCP transport.

Control Channel. This channel transports H.245 protocol messages [3]. It provides signaling communication between H.323 endpoints to control (open and close) Logical Channels and to negotiate and modify media parameters. This channel may also be used for authentication and encryption, according to Rec. H.235 [4]. It uses TCP transport.

Logical Channel. This channel transports RTP/RTCP messages (Section 20.4.1) for media (bearer) flows. It consists of a RTP and a RTCP connection between two H.323 endpoints, set up via H.245 messages on the Control Channel. There can be one bidirectional channel or multiple unidirectional channels in a connection, allowing for asymmetric media communication. This channel uses UDP transport.

21.2.3 H.323 Call-Control Messages and Parameters

H.323 signaling messages are bit-oriented and formats for most of them are specified with the ASN.1 notation (Section 16.2 and [5]), which makes it impractical to describe bit-level structures in concise form. This section describes those messages at a fairly high level and the reader interested in the details is referred to the relevant ITU-T recommendations.

RAS Messages. These messages use the RAS Channel and the most important ones are listed in Table 21.2-2.

ARQ is sent by an endpoint to request, and ACF by the GK to grant, admission to the network for a call.

TABLE 21.2-2 H.225 RAS Messages

Acronym	Message Function	Acronym	Message Function
ACF	Admission confirm	IRR	Information request response
ARQ	Admission request	LCF	Location confirm
BCF	Bandwidth confirm	LRQ	Location request
BRQ	Bandwidth request	RAC	Resource availability confirmation
DCF	Disengage confirm	RAI	Resource availability indication
DRQ	Disengage request	RCF	Registration confirm
IRQ	Information request	RRQ	Registration request

BRQ is sent by an endpoint to request, and BCF by the GK to grant, a change of bandwidth for the call.

DRQ is sent by an endpoint to notify the GK that it is dropping out of the call, and by the GK to an endpoint to force the termination of the call. DCF is the confirmation response.

IRQ is sent by the GK to an endpoint to request a status report, and IRR is sent by the endpoint to respond to the request.

LRQ is sent by an endpoint to request address translation from a GK, and LCF is sent by the GK to respond to the request.

RAI is sent by a GW to inform the GK of available resources, and RAC is sent by the GK as the response. The GK uses the information in its routing decisions.

RRQ is sent by an endpoint to request registration with a GK, and RCF is sent by the GK to respond to the request.

All request messages (ending in RQ, such as ARQ, BRQ, etc.) can also be rejected with a corresponding message ending in RJ (such as ARJ, BRJ, etc., not shown in Table 21.2-2).

The most significant parameters, and the messages they appear in, are listed in Table 21.2-3.

Recommendation H.225 specifies message formats using the ASN.1 notation. Detailed formats are found in [2].

Q.931 Messages. These messages use the call signaling channel and are described in Chapter 10. Not all Q.931 messages are used by H.323: for example, of the messages listed in Table 10.3-1, CONACK, DISC, and RLSE are not used. The deviations from Rec. Q.931 are specified in Rec. H.225 [2].

H.245 Messages. These messages use the control channel and come in four types:

1. *Request* requires an action and a response from the recipient.
2. *Response* is a response to a request.
3. *Command* requires an action but no response from the recipient.
4. *Indication* does not require any action or response.

TABLE 21.2-3 Parameter Fields in RAS Messages

Parameter Name	Message Acronyms													
	ACF	ARQ	BCF	BRQ	GCF	GRQ	IRQ	IRR	LCF	LRQ	RAC	RAI	RCF	RRQ
bandWidth	x	x	x	x				x	x	x				
callModel	x	x						x						
callReference Value		x		x			x							
callSignal Address								x	x				x	x
callType		x		x				x						
destCallSignal Address	x	x												
destinationInfo	x	x							x	x				
gatekeeper Identifier					x	x							x	x
integrityCheck Value	x	x	x	x	x	x	x	x	x	x	x	x	x	x
RAS address					x	x		x	x					x
requestSeqNum	x	x	x	x	x	x	x	x	x	x	x	x	x	x

A selection of H.245 protocol messages is listed in Table 21.2-4. The most significant parameters for request and response messages are listed in Table 21.2-5. The message acronyms in Table 21.2-4 are not official acronyms from Rec. H.245 and are used only for the interpretation of Table 21.2-5.

Most message names are self-explanatory except for two that may need an explanation: *userInputIndication* and *masterSlaveDetermination*. The first is an indication message that carries the equivalent of DTMF digits used in the SCN to convey additional user information after call setup. The second is a set of messages exchanged between two H.323 endpoints to determine which one is in control when both request a resource that can only be assigned to one of them. The determination is based on numbers preassigned to the endpoints for that purpose.

TABLE 21.2-4 H.245 Messages

Type	Message Name	Type	Message Name
Request	CloseLogicalChannel (CLC) MasterSlaveDetermination (MSD) OpenLogicalChannel (OLC) RequestChannelClose (RCC) TerminalCapabilitySet (TCS)	Command	ConferenceCommand EncryptionCommand EndSessionCommand FlowControlCommand SendTerminalCapabilitySet
Response	CloseLogicalChannelAck (CLCA) MasterSlaveDeterminationAck (MSDA) OpenLogicalChannelAck (OLCA) RequestChannelCloseAck (RCCA) TerminalCapabilitySetAck (TCSA)	Indication	MasterSlaveDeterminationRelease OpenLogicalChannelConfirm RequestChannelCloseRelease TerminalCapabilitySetRelease UserInputIndication

TABLE 21.2-5 Parameter Fields in H.245 Messages

	Message Acronyms									
Parameter Name	CLC	CLCA	MSD	MSDA	OLC	OLCA	RCC	RCCA	TCS	TCSA
capabilityDescriptors									x	
capabilityTable									x	
cause										
decision				x						
forwardLogicalChannelNumber		x			x	x	x	x		
forwardLogicalChannelNumbersource	x									
forwardLogicalChannel Parameters					x					
sequenceNumber									x	x
statusDeterminationNumber			x							

Recommendation H.245 specifies messages using the ASN.1 notation. Detailed formats are found in [3].

21.2.4 H.323 Signaling Sequences

The basic signaling sequence for a H.323 call is summarized in Table 21.2-6, which shows how the different protocols under the H.323 umbrella work in concert to build, manage, and release a call.

Call control can be divided into six phases. In the first phase, the calling terminal or gateway uses the RAS protocol (GK admission) to obtain permission from its GK to place the call. Connection control then switches (second phase) to the Q.931 protocol, which sets up the call by building the signaling connection and alerting (ringing) the called party. The called party may also use RAS to obtain permission from its GK (if configured) to accept the call. In the third phase, H.245 is used to handle negotiation of media capabilities between the originating and the terminating endpoints, and to pass that information to RTP/RTCP. In the fourth phase, the media streams are activated by RTP/RTCP. The fifth phase is the media connection and H.245 may be used during this phase to modify media parameters or, together with Q.931, to manage supplementary services. In the final phase, the Q.931 protocol is used to close the signaling channels, and the RAS protocol is used to release the gatekeeper(s) if they stayed on during the call.

In addition to the six phases just described, a terminal or a gateway may use the RAS protocol to find its gatekeeper and to register with it, before starting a call. Gatekeeper discovery and registration must be done when initializing or reconfiguring a terminal or a GW; but once done it remains valid for subsequent calls.

Many call scenarios are possible within the general framework of Table 21.2-6, based on the following factors:

(a) Direct call signaling between endpoints or GK-routed call signaling

(b) Control channel routed directly between endpoints or control channel routed through GKs

TABLE 21.2-6 H.323 Basic Call Sequence

	Phase	Protocol	Function	Notes
GK setup	GK discovery	H.225/RAS	Endpoint finds where its GK is	1
	GK registration	H.225/RAS	Endpoint registers with its GK	1
Call control	GK admission	H.225/RAS	Endpoint obtains permission to call	2
	Call setup	H.225/Q.931	Establishing connection between endpoints	
	Media capability exchange	H.245	Endpoints negotiate media parameters	
	Media communication	RTP/RTCP	Media communication through logical channels	
	Call services	H.245, Q.931	Changes to media parameters and/or supplementary services may take place	
	Call termination	H.225 (Q.931, RAS), H.245	Connection is ended	3

Notes:

1. These two phases take place only if an endpoint needs to register with its GK.
2. This phase takes place for every call except when no GK is present.
3. Since the call signaling channel may not need to stay up for the whole duration of the connection, the Q.931 termination sequence may happen earlier in the call, although obviously after call setup. The decision is made by the GK.

(c) Point-to-point connection or multipoint connection (conference)

(d) Voice, video, data, fax

Items (a) and (b) above refer to the fact that a H.323 network may be configured with or without gatekeepers and that when a GK is present, the GK may be configured only for admission control and address translation or it may remain involved in the call to relay signaling messages to/from endpoints.

Figure 21.2-5 shows an example of a message sequence for a point-to-point call where both the calling and the called party are in the same zone and where the GK allows direct call signaling between terminals. The sequence in Fig. 21.2-5 is typical of full-featured H.323 terminals. SET terminals use a simplified *fast connect* procedure that exchanges media parameters inside Q.931 call-setup messages.

Feature Control of Supplementary Service. Three ways of handling supplementary services are supported by H.323:

1. Functional feature control, specified in Rec. H.450 [6]
2. Stimulus feature control, specified in Annex L of Rec. H.323
3. HTTP feature control, specified in Annex K of Rec. H.323

Recommendation H.450 follows the traditional functional approach, where the control functions for each service are specified individually.

Annex L of Rec. H.323 defines a stimulus protocol for feature control, which can be used to relay key (pushbutton) operations from the terminal to a feature server and to display information from the feature server on the terminal (see Section 6.1 for a definition of stimulus protocols).

Annex K of Rec. H.323 describes a mechanism by which the Hypertext Transfer Protocol (HTTP) [7] is used on a HTTP control channel to give an H.323 terminal access to customized web pages. The URL of the web page is sent to the terminal by the GK in a RAS message (RCF). Using the URL, the terminal

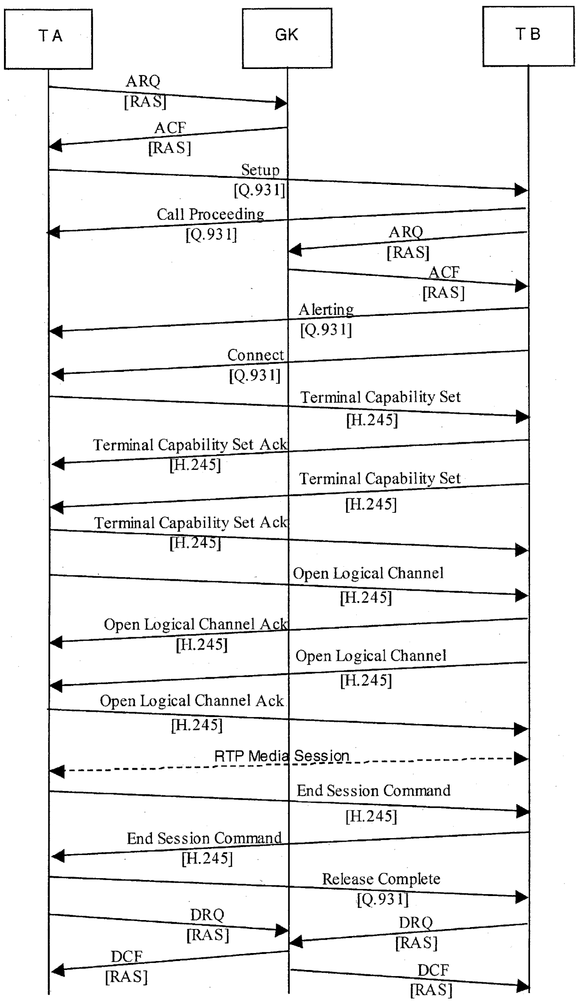


Figure 21.2-5. H.323 call flow example.

sets up the HTTP connection to the web page, from which it downloads feature control information.

21.3 THE SESSION INITIATION PROTOCOL (SIP)

21.3.1 Overview

SIP is an IETF standard, the first draft of which was published in 1996. It became a Proposed Standard in 1999 as RFC 2543. The current version is RFC 3261, which is backward compatible with RFC 2543 [8–10]. SIP was based from the start on TCP/IP and Internet concepts, to manage multimedia connections (two-party and multiparty) in IP-based packet networks. SIP documents use the term *session*, instead of *call,* for media communication, to highlight the multimedia aspects of the protocol.

Like H.323, SIP is a peer-to-peer protocol based on the direct-routed model (distributed control) and is typically deployed for signaling between SIP endpoints (user agents and MGCs). It can also be used for communication between SIP endpoints and supporting servers that interact with the endpoints to provide routing and access control.

SIP is not an all-inclusive, vertically integrated protocol suite like H.323: it supports only two basic functions for multimedia session control, namely,

- Determination of user availability, location, and capabilities
- Session setup and management

For functions beyond the ones above, SIP systems are expected to use other protocols, based on choices negotiated between SIP signaling points. In particular, the media path and its establishment are outside the scope of SIP. SIP is used by the origination and the destination to learn each other's network address and port; it then enables the negotiation for the protocols, services, and formats to be used for the media connection. An optional portion of a SIP message (*message body*—Section 21.3.4) is passed on by SIP unmodified, offering a free, direct exchange of media parameters between calling and called party, and an open platform for new services. SIP's limited scope and "open" approach is both simple and flexible.

Message formats and syntax are similar to HTTP [7]; as a result, SIP offers smooth interoperability with Internet protocols such as RTP, SDP, and the Dynamic Host Configuration Protocol (DHCP—an application that provides IP addressing assistance to users).

SIP is extendable by design: SIP signaling entities may ignore items that are not recognized but do not refuse requests containing them, resulting in backward compatibility of protocol extensions.

For Layer 4, UDP and TCP support is mandatory, but other transport protocols, such as the Stream Control Transmission Protocol (SCTP), are allowed as options.

Since SIP can be and is used in networks connected to the Internet, one area of concern is unauthorized access. That concern is addressed by procedures for authentication (with methods similar to HTTP's) and encryption (via existing Internet protocols).

21.3.2 SIP Components

SIP uses a client-server, request-response mode of operation and all the components of a SIP network, called *servers*, can perform both client and server roles. Five types of servers may be found in a SIP network:

1. User agent (UA)
2. Proxy server
3. Redirect server
4. Registrar
5. Location service

Consistent with the distributed control model, SIP terminals (user agents) contain call-control logic and can communicate with each other directly. Proxy, redirect, registrar, and location service servers play supporting roles.

User Agent (UA). This server originates or services a request on behalf of a user. A SIP user agent can be of two kinds: *user agent client* (UAC) and *user agent server* (UAS). The former acts on behalf of a caller, the latter on behalf of a callee. UAC and UAS typically are two roles played by the same application program. An example is a "softphone" application running on a personal computer, which acts as a UAC when originating a call and as a UAS when receiving a call. A media gateway controller is the UA for each SCN customer connected to the SIP network through the MGs it controls.

Proxy Server. This server is an intermediary entity that helps a UA with routing and manages UA access to network resources through authentication and authorization. Requests from a UA may be serviced directly or forwarded to other servers, sometimes after being partially rewritten. As the name implies, this server acts as a proxy for a UA and is similar in function to an H.323 gatekeeper. Depending on the *transaction* (see next section), proxy servers can operate in either stateful or stateless mode: that is, they may base their next action on memory of previous events or simply forward a message and forget. Stateful and stateless refer to the state of the transaction, not to call state. Stateless mode is faster but more limited in functionality.

Redirect Server. This server returns one or more addresses upon receiving a request from a client (a UA or a proxy). It cannot initiate requests; its only function is to respond to a client and provide new addresses to *redirect* the call. After servicing a request a redirect server returns control of the call to the source of the request. In H.323 networks the gatekeeper performs the function of the redirect server.

Registrar. This server receives a user's current location by servicing a REGISTER request from a UAC and then records that information in a location service (see below). The registrar's only function is to receive and record location information, not to provide it to clients; that role is played by the location service. The registrar function is typically implemented in a server that also hosts proxy or redirect functions. The interface between a registrar and a location service is outside the scope of the SIP specification.

Location Service. This server provides a called party's possible locations to a requesting redirect or proxy server. A location service may obtain location information from a registrar or from other sources. The interface between a location service and registrars, proxy, or redirect servers is outside the scope of the SIP specification.

21.3.3 SIP Sessions

SIP sessions typically involve four components: two user agents (one for the source, the other for the destination) and two proxy servers (one proxying for the source and one for the destination). The standard configuration of a SIP connection (*SIP trapezoid)* is shown in Fig. 21.3-1.

Two forms of message exchanges are used for signaling communication between SIP entities:

- Dialog
- Transaction

A dialog is a logical connection between two user agents and was called a *call leg* in the original version of SIP. A dialog starts when a called UA accepts and starts processing a request for a connection from a calling UA. In each UA a dialog is identified by a dialog ID, consisting of three parameters: Call-ID, local tag, and remote tag. The Call-ID is globally unique; that is, both UAs involved in a dialog have the same Call-ID. The local tag is assigned separately by each UA; the remote tag has the value of the local tag of the other UA (see next section for more details on tags).

A transaction is the basic building block of SIP signaling. It consists of an exchange of messages between a client and a server, encompassing a request and its responses. A call setup typically requires multiple transactions, depending on

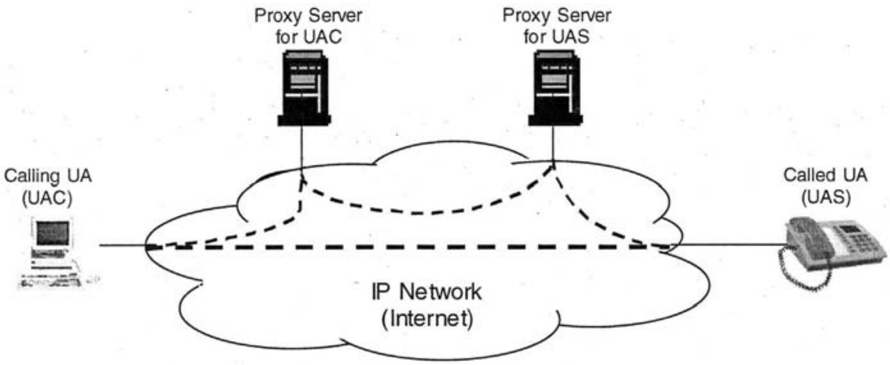


Figure 21.3-1. Entities and connections in a point-to-point SIP session setup.

the number of servers involved. A similar construct, also called transaction, is found in TCAP (Chapter 16).

21.3.4 SIP Messages

SIP messages can be divided into two categories:

1. Request, from a client to a server
2. Response, from a server to a client

Request messages have six message types (Table 21.3-1), identified by the value of the *Method* parameter. Responses do not have message types of their own, although the value of the *Status Code* parameter (Table 21.3-2) could be viewed as a kind of message type.

Message Format. Messages consist of lines of text, each terminated by Carrier Return and Line Feed (CRLF). The small increase in bandwidth due to the use of text instead of bits is not significant given today's bandwidth costs and availability.

TABLE 21.3-1 SIP Request-Message Types (Method Parameter Values)

Message	Function
INVITE	Initiates sessions
ACK	Confirms session establishment
OPTIONS	Requests information about capabilities
BYE	Terminates an established session
CANCEL	Cancels a pending session
REGISTER	Registers a user's current network location at which it can be reached (for the personal mobility function)

TABLE 21.3-2 Status Codes and Reason Phrases (SIP Response Messages)

Status Code	Reason Phrase	Status Code	Reason Phrase
100	Trying	410	Gone
180	Ringing	416	Unsupported URI scheme
200	OK	483	Too many hops
400	Bad request	484	Address incomplete
401	Unauthorized	486	Busy here
404	Not found	503	Service unavailable
408	Request timeout	504	Server timeout
415	Unsupported media type	600	Busy everywhere

The general message format is the same for both requests and responses and is similar to the format of HTTP messages:

1. Start line
2. Header lines (headers)
3. Empty line (with CRLF)
4. Body (optional)

The first set of text lines, comprising items 1, 2, and 3 above, follows a format precisely defined in the SIP specification and provides the information needed for routing and interpreting the message. The precisely formatted part is followed (separated by a blank line ending with CRLF) by an optional set of lines that contain the *message body*, whose format is unspecified.

The start line defines the message as a request or a response. The header fields provide additional information that cannot fit into the start line and control the interpretation of the message body. The order in which multiple headers appear in a message is not significant. Some headers are unchanged end-to-end, and some may be modified by proxies.

The most important function of the message body is to carry *session descriptions*, which are used by the parties involved in a session to agree on media stream characteristics. The default protocol for session descriptions is SDP (Section 20.4.3) [11]. The message body may also be used for human-readable text, to provide call-related information to users, as well as for other data needed to implement services.

SIP uses the ABNF notation [12] for message formats, whose complexity makes its use impractical for describing message formats in concise form. In this chapter, therefore, we describe message formats using a layperson's notation, where words in *italics* represent the names of parameters (replaced by values in actual messages), and where items in **bold** (and not in italics) appear literally as they would in actual messages. **SP** denotes a space character and **CRLF** denotes a carriage return followed by line feed.

The Request message format is as follows:

1. Request line (start line):

   Method **SP** *Request-URI* **SP** *SIP-Version* **CRLF**

2. Header lines:

   *header-name***:** *header-value* **CRLF**

3. Empty line ending with CRLF:

   CRLF

4. Message body lines:

   message-body

The Method parameter indicates the type and function of the request and can have the six values shown in Table 21.3-1. The Request-URI is a URI (more on it below) that indicates the desired destination of the request.

The Response message format is as follows:

1. Status line (start line):

   SIP-version **SP** *Status-Code* **SP** *Reason-Phrase* **CRLF**

2. Header lines:

   *header-name***:** *header-value* **CRLF**

3. Blank line with CRLF:

   CRLF

4. Message body lines:

   message-body

The Status-Code is a number that denotes the result of processing a request, and the Reason-Phrase is a textual description of that result. Examples of Status Code and Reason Phrase values are shown in Table 21.3-2.

Headers. Some headers are mandatory and some are optional: the rules governing under what circumstances a particular header may or may not be used are based on the type of message (Request or Response), Status Codes, or Method.

The following headers are mandatory in request messages: To, From, Cseq, Call-ID, Max-Forwards, and Via.

Response messages normally repeat the mandatory headers of the request, except for Max-Forwards, which is used only in requests. Error responses do not repeat the mandatory headers.

Most headers contain parameters, separated by semicolons. Individual headers can be split across multiple lines using a SP character or horizontal tab (HT). Below is the syntax for all the mandatory headers in request messages. The examples for each header are from RFC 3261 [8].

To. This header indicates the target or desired recipient of a message and is echoed back in response messages.

$$\textbf{To}{:}\textit{SIP URI}\textbf{;}\ \textbf{tag} = \textit{parameter}$$

A display name may precede the URI, which must then be put in angle brackets. The tag parameter is not present if a dialog is not yet established. For example:

To: Carol ⟨sip:carol@chicago.com⟩

To: Alice ⟨sip:alice@atlanta.com⟩;tag = 1928301774

To: sip:+12125551212@server.phone2net.com

From. This header indicates the logical identity of the initiator of the request. Responses echo the information back.

$$\textbf{From}{:}\ \textit{SIP URI}\textbf{;}\ \textbf{tag} = \textit{parameter}$$

A display name may precede the URI, which must then be put in angle brackets. For example:

From: Bob ⟨sip: bob@biloxi.com⟩;tag = 456248
From: sip:+12125551212@ phone2net.com;tag = 887s

Cseq. This header stands for "command sequence" and contains a decimal integer that is incremented by one for every successive request within a given transaction. Responses echo the number back.

$$\textbf{Cseq:}\ \textit{Sequence-Number Method}$$

For example:

Cseq: 4711 INVITE

Call-ID. This header is an alphanumeric identifier that links together a group of messages to form a dialog. It is assigned by a UAC when starting a dialog.

$$\textbf{Call-ID:}\ \textit{call-id@host}$$

For example:

Call-ID: a84b4c76e66710@ pc33.atlanta.com

The @*host* portion is optional.

Max-Forwards. This header is a counter that typically starts at 70 and is decremented by one at each "hop." If it reaches 0 the request is rejected. It avoids wasting resources when a request message ends in a routing loop.

Max-Forwards: *hop-number*

For example:

Max-Forwards: 6

Via. This header indicates the path followed by the request through the network. The original client inserts the first Via header, then each proxy server adds its own (except in the case of redirection), so multiple Via headers may be present in the same message. This information indicates to recipients the path that responses must follow:

Via:*protocol/version/transport host***:***port***;** *Via-parameter comment*

For example:

Via: SIP/2.0/UDP erlang.bell-telephone.com:5060;branch=z9hG4bK87asdks7
Via: SIP/2.0/UDP192.0.2.1: 5060 ;received=192.0.2.207;branch=z9hG4bK77asjd

Parameters. Two important parameters used in SIP headers are SIP URI and tag. They are discussed below. The examples are from RFC 3261 [8].

SIP URI. This parameter is similar to Internet URIs (Section 20.2.5) [13] and identifies source, destination, or redirection addresses in SIP messages. Its typical form is

sip :*user@host*

The complete form (with all the optional fields) is

sip:*user***:***password@host***:***port***;***URI-parameters?headers*

User. This parameter can be a telephone number, in which case it uses the format for the "tel URL" specified in RFC 2806 [14], characterized by a " + " in front of the number.

An additional type of URI, called SIPS URI, is used to indicate that the owner of an address wants to be contacted on secure channels. SIPS URIs have the same

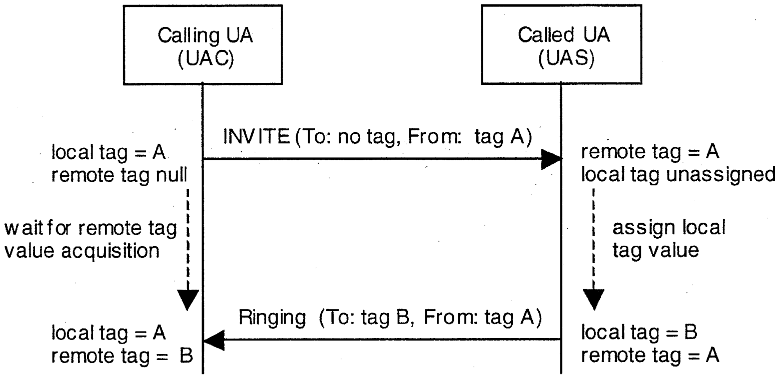


Figure 21.3-2. Tag value assignment and acquisition.

format as SIP URIs, except for "sips" instead of "sip." For examples:

sip:bob@biloxi.com

sips:1212@gateway.com

SIP also allows the use of Internet URIs in lieu of SIP URIs.

Tag. This is a parameter in the "To" and "From" headers, consisting of a randomly generated but unique character string. Two tag values, local and remote, are used by each UA, and together with the Call-ID they identify a dialog (Dialog ID). A UA assigns its own local tag value and acquires the remote tag value from the remote UA.

In a request message, the "To" header carries the remote tag; prior to its acquisition, the field is omitted. The "From" header carries the local tag. In a response message, the roles are reversed: "To" carries the local tag and "From" carries the remote tag.

The process for assigning the local tag and for acquiring the remote tag is shown in Fig. 21.3-2. The calling UA inserts its local value (tag=A) in the "From" header of the INVITE message, starting the dialog. Upon receipt of the INVITE, the called UA acquires A as its remote tag value and assigns its own local value (tag=B). Tag = B is inserted in the "To" header of the first response message (e.g., Ringing) sent to the calling UA, which acquires it as its remote tag. At this point the dialog is established (the Call-ID was assigned by UA and sent in the INVITE). In Fig. 21.3-2 the INVITE and Ringing messages may be sent directly or relayed by proxy servers (not shown).

21.3.5 SIP Call Sequences

A SIP session is conceptually very simple, as shown in the example of Fig. 21.3-3, taken from RFC 3261 [8]. The example is based on a point-to-point session

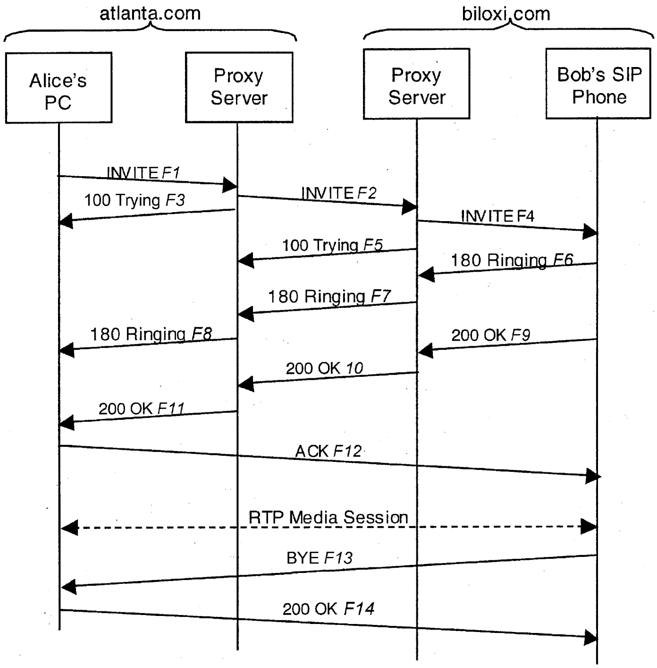


Figure 21.3-3. Example of a SIP call sequence. (From RFC 3261 [8].)

(a regular voice call) between Alice, equipped with a SIP-enabled PC, and Bob, equipped with a SIP phone; Alice originates the session. In Fig. 21.3-3, messages are labeled F1 through F14 for reference in the description below, and numbers like 100, 180, and 200 are Status Code parameters in the response messages.

The session starts with an invitation from the caller UAC to the called UAS to join the session. The called UAS is identified by a SIP URI, and the invitation takes the form of an INVITE request (F1). The INVITE may contain a *session description* in the message body, describing the media characteristics that the caller UAC supports. The calling party's UAC is configured to send the INVITE request to a proxy server (*outbound proxy*) acting on behalf of the caller. The outbound proxy can be preassigned or the caller's UAC may use DHCP to determine its address. The outbound proxy queries a DNS server to locate the proxy server for the called party and forwards the INVITE to it (F2, F3). The called party's proxy contacts a location service to find the called party's current whereabouts and then forwards the INVITE to its final destination (F4, F5). The UAS of the called party accepts the invitation with a response message, which may include a media description indicating acceptance of the media parameters (F6–F11). The caller's UAC confirms the session by sending an *acknowledge* (ACK) request to the called party's UAS, which completes the initial transaction (F12).

At this point the caller and called party start the media session according to the agreed-to parameters exchanged in the message bodies.

A call ends when the UAC for either party sends a BYE request, after which media communication is terminated (F13, F14). Invitations that are not yet completed may be canceled using a CANCEL request.

The media connection is independent of the signaling process and takes place using protocols and formats that are negotiated by the exchange of information in the body portion of the signaling messages.

A variation of this sequence is when the redirect mode of operation is used instead of the proxy mode. In that case, the calling UAC sends the INVITE to a redirect server, which, after querying DNS or a location service responds to the requesting UAC with routing information in the Contact header field. The Caller's UAC then reissues the INVITE request using the contact information, after which the call proceeds as in the previous case. The main advantage of the redirect mode of operation is server efficiency, since it reduces the overall number of messages exchanged for session setup.

SDP and Media Parameter Negotiation. When SDP is used, the response to a request containing media descriptions repeats the "m=" and "a=" lines from the request and indicates acceptance of a media stream by replacing the port numbers of the sender with the port numbers of the responder. Rejection of a media stream is indicated by a 0 [zero] port number. The response cannot contain any new parameters or new values, only acceptance or rejection of individually offered streams. The called party could also reject the whole media proposal (with an appropriate status code in the response), in which case the calling party has to issue a new INVITE.

21.3.6 Examples of SIP Messages

Following are two examples of SIP messages in the SIP session of Fig. 21.3-3, based on the following attributes:

- Alice's SIP URI: sip:alice@atlanta.com
- Alice's personal computer's domain name: pc33.Atlanta.com
- Alice's personal computer's IP address: 192.0.2.1
- Alice's outgoing proxy's domain name: bigbox3.site3.atlanta.com
- Alice's outgoing proxy's IP address: 192.0.2.2
- Bob's SIP URI: sip:bob@biloxi.com
- Bob's proxy's domain name: server10.biloxi.com
- Bob's proxy's IP address: 192.0.2.3
- Bob's phone's IP address: 192.0.2.4 (a domain name is not used in this example)

The two message examples are from RFC 3261 [8].

F1—INVITE Request From Alice → atlanta.com proxy

INVITE sip:bob@biloxi.com SIP/2.0
Via: SIP/2.0/UDP pc33.atlanta.com;branch=z9hG4bKnashds8
Max-Forwards: 70
To: Bob ⟨sip:bob@biloxi.com⟩
From: Alice ⟨sip:alice@atlanta.com⟩;tag=1928301774
Call-ID: a84b4c76e66710@pc33.atlanta.com
CSeq: 314159 INVITE
Contact: ⟨sip:alice@pc33.atlanta.com⟩
Content-Type: application/sdp
Content-Length: 142

F9—Response 200 OK From Bob→biloxi.com proxy

SIP/2.0 200 OK Via: SIP/2.0/UDP server10.biloxi.com;branch=z9hG4bK4b43c2ff8.1;received=192.0.2.3
Via: SIP/2.0/UDP bigbox3.site3.atlanta.com;branch=z9hG4bK77ef4c2312983.1;received=192.0.2.2
Via: SIP/2.0/UDP pc33.atlanta.com;branch=z9hG4bKnashds8;received=192.0.2.1
To: Bob ⟨sip:bob@biloxi.com⟩;tag=a6c85cf
From: Alice ⟨sip:alice@atlanta.com⟩tag=1928301774
Call-ID: a84b4c76e66710@pc33.atlanta.com
CSeq: 314159 INVITE
Contact: ⟨sip:bob@192.0.2.4⟩
Content-Type: application/sdp
Content-Length: 131

21.4 THE GATEWAY CONTROL PROTOCOL

The Gateway Control Protocol, formerly known as Megaco, is a signaling standard issued jointly by IETF and ITU-T. The IETF standard is RFC 3525 [15], and the ITU-T standard is the H.248 series of recommendations [16]. The IETF and ITU-T requirements are the same, so in this chapter we use "Gateway Control Protocol" and "H.248 protocol" interchangeably. This protocol is used by media gateway controllers (MGCs) to control media gateways (MGs). It is independent of the lower layers but, if used "over" IP, the MGC must support both TCP and UDP, and the MG must support either TCP or UDP.

The Gateway Control Protocol evolved from an earlier protocol, the Media Gateway Control Protocol (MGCP), which was based on similar principles and is still in use [17].

21.4.1 Network Model

The network model for the Gateway Control Protocol is shown in Fig. 21.4-1. It is designed primarily to support SCN interfaces (lines and trunks) that connect to a packet network via a media gateway controlled by a media gateway controller. The approach is typical of the server-routed model (centralized control): call-control logic is located in the MGC (network), and MGs (peripherals) have little intelligence.

A MGC uses this protocol to communicate with a MG, while communicating with other MGCs (or UAs) via peer-to-peer protocols like H.323, SIP, or BICC. A MGC and the MGs it controls are the only units needed to handle lines and channel-associated signaling (CAS) trunks to/from the SCN. Common-channel signaling trunks, on the other hand, require the additional presence of a signaling gateway. The SG is needed to interface with the signaling data links from/to the SCN. The SG can be integrated into the same physical unit as the MGC or can be a separate physical unit, in which case it communicates with the MGC via a protocol like SIGTRAN, described later in this chapter.

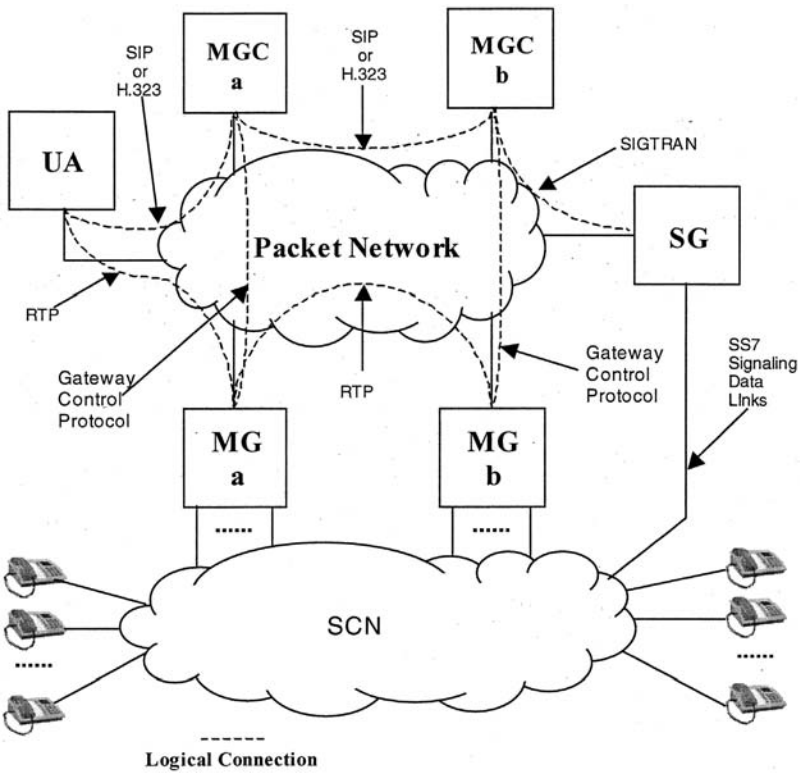


Figure 21.4-1. Network model for the Gateway Control Protocol.

21.4.2 Signaling Protocol

The H.248 protocol is a master–slave protocol, with the MG under the direct control of the MGC. It is described in terms of two logical entities (called *abstractions*) located in the MG:

- *Termination*—the source or the destination of one or more user or control (signaling) data streams
- *Context*—an association between a set of terminations

Calls are established by adding terminations to contexts and are cleared by deleting terminations from contexts. A voice connection in which an analog interface is connected to a packet channel is described as a context with two terminations: the analog termination and the packet termination (Fig. 21.4-2(*a*)). A special context, called the *null context,* does not correspond to any connection but is used to park physical terminations when they are idle (Fig. 21.4-2(*b*)). A line on hold is described as a context with only one termination (Fig. 21.4-2(*c*)).

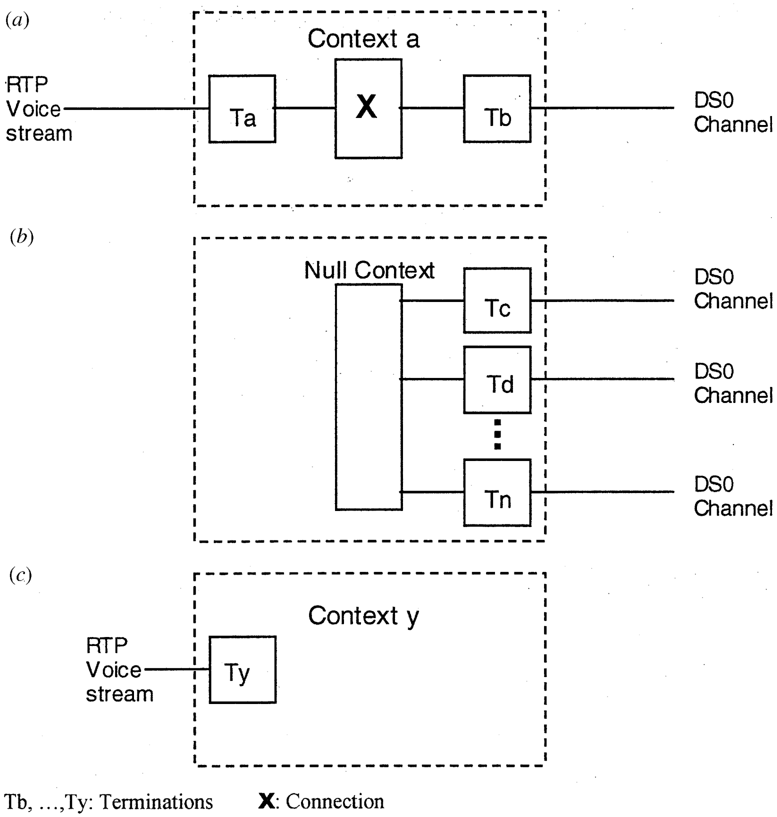


Figure 21.4-2. Terminations and contexts for the Gateway Control Protocol.

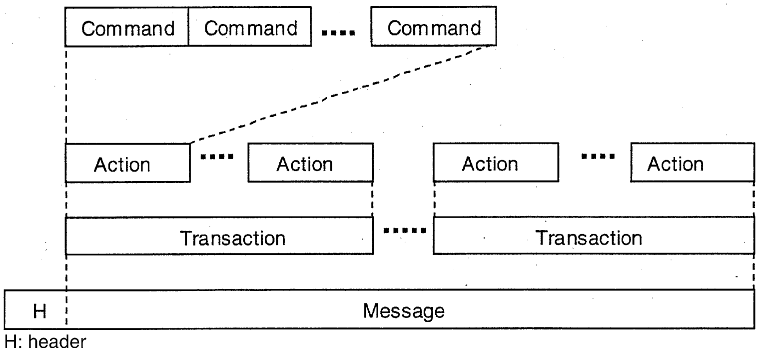


Figure 21.4-3. Message structure for the Gateway Control Protocol.

Terminations and contexts are controlled by *commands*. A group of commands related to a context constitutes an *action*. One or more related actions constitute a *transaction*. Transactions are managed by means of *messages* (Fig. 21.4-3). A message is just a vessel for sending and receiving transaction commands, since the protocol does not have any provision for acknowledgment of messages. Transactions (see later in the section) have a request–response mechanism.

Terminations. These are abstractions representing the entry or exit point of one or more *streams* into or from a MG. Streams are unidirectional or bidirectional logical channels that carry user or control data. For a normal call a termination represents one bidirectional voice stream; for a multimedia call it represents multiple streams, such as voice and video. A termination that represents a packet channel is called *ephemeral* and is created and deleted as needed on a call-by-call basis. A termination that represents a physical port, such as a TDM trunk, is semipermanent: it is created when provisioned and lasts as long as provisioned. Physical terminations are moved in and out of the null context, which corresponds to their idle state (Fig. 21.4-2(*b*)).

Terminations are identified by the TerminationID parameter, chosen by the MG. Groups of terminations can be addressed via special character strings called *wildcards*; the *root* termination can be used to addressed all terminations in a MG.

Parameters that define a termination's characteristics are called *properties* and are grouped into sets called *descriptors*. Descriptors can be nested, that is, a descriptor can be composed of descriptors. Examples are shown in Table 21.4-1.

Events, Signals, and Statistics descriptors, together with individual properties, can be grouped into *packages*. A termination is characterized by a set of packages.

Some packages are defined in the basic standard (Table 21.4-2 shows examples) but they can also be added later outside the current version of the standard, thus providing a tool for protocol extension. Packages introduced as extensions can be defined by the IETF (as RFCs) or by ITU-T (as recommendations in the H.248 series).

Contexts. These represent connections between terminations. They are created by the MG and are identified by the ContextID parameter (an integer).

TABLE 21.4-1 Gateway Control Protocol Descriptors

Descriptor	Function
Media	List of media stream specifications
Stream	Describes an individual stream in a termination
LocalControl	Properties of MG−MGC connection
Local	Properties of media from far end of connection
Remote	Properties of media to far end of connection
Events	Events to be detected and action required by MG
Signals	Signals applied by MG to a termination
DigitMap	Dialing pattern used by MG to detect digits in blocks rather than individually
ServiceChange	Describes service change
Statistics	Statistics accumulated by MG for a termination
Packages	List of packages for termination
ObservedEvents	Report events to MGC
Topology	Flow direction in a context
Error	Error code

TABLE 21.4-2 Packages for the Gateway Control Protocol

Package	Function
Base root	Properties applying to the whole MG
Tone generator	Audio tones that can be generated
Tone detection	Characteristics of detectable audio tones
Basic DTMF generator	DTMF tones that can be generated
DTMF detection	Detectable DTMF tones
Analog line supervision	Events and Signals for analog lines
Network	Properties of network terminations (e.g., digital channel)
RTP	Properties of RTP streams

Commands. These are used to control terminations and contexts. They are listed in Table 21.4-3. Commands use descriptors as parameters.

Transactions. These are a set of actions related to the same call, and the commands they contain are executed sequentially. Transactions, on the other hand, may be executed out of order.

Transactions have two components:

1. TransactionRequest, used to request the recipient to execute a series of commands
2. TransactionReply, used by the recipient of a request to respond to the commands in the request

In the TransactionReply, the recipient of a request returns the values that the request's parameters acquire after the execution of the commands. The recipient

TABLE 21.4-3 Gateway Control Protocol Commands

Commands	Function	MGC ↔ MG
Add	Add termination to context	→
Modify	Modify properties of termination	→
Subtract	Remove termination from context	→
Move	Move termination to another context	→
AuditValue	Return current state of properties and statistics of termination	→
AuditCapabilities	Return supported values of properties and statistics of termination	→
Notify	Notify MGC of events in MG	←
ServiceChange	Notify MGC of physical termination service status change, order	→
	MG to take terminations in/out of service	←

can also respond with a TransactionPending, to notify the sender that the commands are being processed.

The format of a TransactionRequest is as follows (from RFC 2535 [15]):

TransactionRequest (TransactionID {
ContextID {Command ... Command},
...
ContextID {Command ... Command }})

The format of a TransactionReply is as follows:

TransactionReply(TransactionID {
ContextID {Response...Response},
...
ContextID{Response...Response}})

The value of the TransactionID is assigned by the originator of the request, and the same value is used in the request and in the reply.

Messages. Messages are used to send one or more transactions and consist of a header and a message body. The header contains the *message ID* (MID) and optional authentication information. The MID consists of the domain name or IP address of the sender. The message body contains transactions.

Due to the complexity of the encoding, we refer the reader to Annex A and Annex B of [15] for the detailed message formats.

21.4.3 Call Sequence Example

An example of a call sequence is shown in Fig. 21.4-4. In the example, the call originates from an analog terminal Ta connected to MGa via an analog subscriber line, and terminates at a TDM digital trunk connected at MGb. Both MGs are controlled by the same MGC. The digital trunk on MGb uses SS7 signaling (Chapter 7) and the MGC communicates with a signaling gateway (SG), using an interface and messages that are not shown (see SIGTRAN in the next section). When the called party answers, the MGC is notified of the event by an ISUP message (ANM) via the SG, which causes the MGC to issue a Modify command to MGa to connect the voice path.

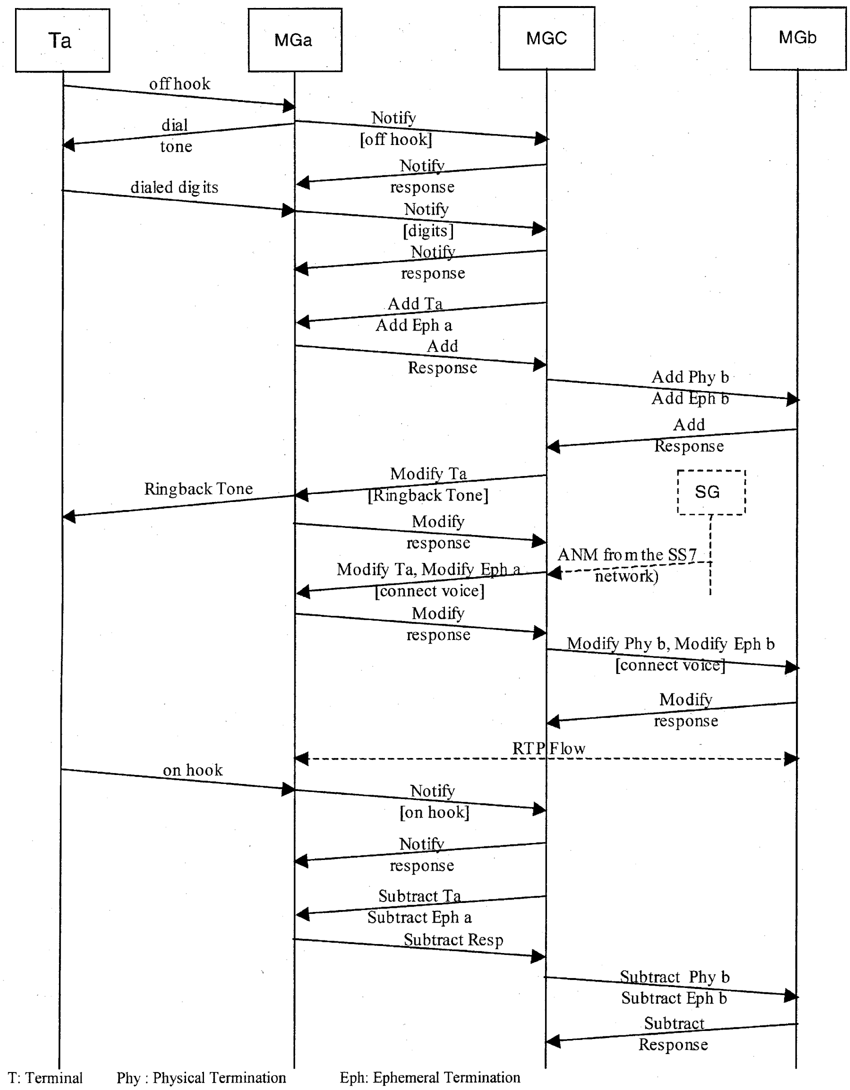


Figure 21.4-4. Example of call sequence for the Gateway Control Protocol.

21.5 THE SIGNALING TRANSPORT (SIGTRAN) PROTOCOLS

21.5.1 Overview

SIGTRAN is a suite of protocols for transporting SCN signaling messages (common channel) inside an IP network. The protocols are issued by the IETF [18]. The typical use of SIGTRAN is between a signaling gateway and a media gateway controller, or between a SG and an IP-based application server that hosts a centralized database such as a SCP.

The components of a SIGTRAN network (SGs, MGCs, and ASs) are called *IP signaling points* (IPSPs), since SIGTRAN uses the IP protocol at Layer 3. An IP entity that processes SCN signaling messages is called an *application server process* (ASP). MGCs and centralized databases (called applications servers in Chapter 20) are examples of ASPs. A set of redundant ASPs working in active-backup mode is called an application server (AS) in SIGTRAN RFCs, but in the rest of the chapter we will use the term as defined in Chapter 20.

Straightforward transport of SCN signaling from the SCN-to-IP boundary (such as a SG) to its processing point in the IP network is called *backhauling*. With backhauling the signaling entities in two IPSPs communicating via SIGTRAN are not aware they are running in two separate nodes.

Three basic network configurations are supported by SIGTRAN (Fig. 21.5-1): the first (a) is backhauling of SCN signaling messages between a SG and the ASP that processes them (MGC/AS); the second (b) is transport of SCN signaling messages between two signaling gateways, when SCN signaling transits through the IP network but is originated and terminated in the SCN; the third (c) is when SIGTRAN signaling is used between IP entities, with no connection to the SCN.

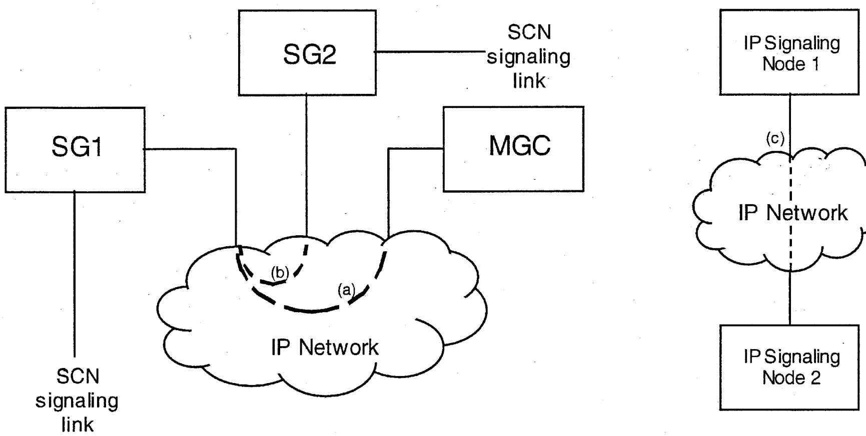


Figure 21.5-1. SIGTRAN network model.

SCN Signaling Applications						
IUA	V5UA	M2UA	M2PA	M3UA	SUA	Adaptation Layer
SCTP						Transport Layer
IP						Network Layer
Data Link and Physical Layers						

Figure 21.5-2. SIGTRAN protocol stack.

The SIGTRAN protocol stack, shown in Fig. 21.5-2, is structured according to three layers:

1. Network Layer: IP protocol.
2. Transport Layer:SCTP (Section 20.4.2) is the recommended protocol, but TCP or UDP are also allowed.
3. Adaptation Layer: it is composed of several modules, each supporting a different SCN protocol.

21.5.2 SIGTRAN Adaptation Layer

The SIGTRAN Adaptation Layer modules and the SCN protocols they support are as follows:

- IUA, supporting ISDN (Q.921—Chapter 10).
- V5UA, supporting V5.2 (LAPV5—Chapter 6)
- M2UA, supporting SS7 MTP2 (Chapter 8)
- M2PA, supporting SS7 MTP2 (Chapter 8)
- M3UA, supporting SS7 MTP3 (Chapter 8)
- SUA, supporting SS7 SCCP (Chapter 15)

Adaptation Layer modules perform two basic functions:

1. *Layer management*, which manages the IP/SCTP connection between signaling points
2. *Adaptation function*, which transfers SCN messages between signaling points in the IP network

When SCTP is the L4 protocol, SIGTRAN is a SCTP user and Adaptation Layer modules are SIGTRAN ULPs (Section 20.4.2).

Adaptation Layer modules send primitives to the SCTP layer, such as Initialize, Associate, Send, and Shutdown. The SCTP layer sends *notifications*, such as Data Arrive, Communication Up, and Shutdown Complete.

Redundancy. Virtually all Adaptation Layer modules support application server process redundancy, for reliability and scalability. Redundancy consists of a network configuration where a SG has connections to more than one ASP, with one connection always active. Redundancy is in addition to the network reliability provided by SCTP multihoming, which protects against faults in the network rather than faults or congestion in ASPs.

Messages. Adaptation Layer modules in separate nodes communicate with each other via messages. Messages are grouped into *message classes* (Table 21.5-1). Some classes are used by more than one module while others are module-specific. All Adaptation Layer messages have a common header, shown in Fig. 21.5-3, followed by the message body, made of optional variable-length parameters (Fig. 21.5-4). Some of the modules also insert a module-specific header between the common header and the variable-length parameters, as discussed in the next sections.

21.5.3 ISDN User Adaptation Protocol (IUA)

This Adaptation Layer module supports both basic and primary rate ISDN interfaces on the SCN side and provides backhauling of Q.931 messages between a SG and an

TABLE 21.5-1 Message Classes for Adaptation Layer Modules

Class Code	Class Name	Acronym	M2UA	M3UA	SUA	IUA	M2PA	V5UA
0	Management messages	MGMT	X	X	X	X		X
1	Transfer messages			X				
2	SS7 signaling network management messages	SSNM		X	X			
3	ASP state maintenance messages	ASPSM	X	X	X	X		X
4	ASP traffic maintenance messages	ASPTM	X	X	X	X		X
5	Q.921/Q.931 boundary primitives transport messages	QPTM				X		
6	MTP2 user adaptation messages	MAUP	X					
7	Connectionless messages	CL			X			
8	Connection-oriented messages	CO			X			
9	Routing key management messages	RKM		X	X			
10	Interface identifier management messages	IIM	X					
11	M2PA messages						X	
12–13	Reserved by the IETF							
14	V5 boundary primitives transport messages	V5PTM						X
15–127	Reserved by the IETF							
128–255	Reserved for IETF-defined message class extensions							

Note: ASP, application server process.

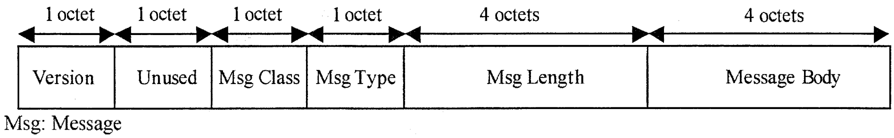


Figure 21.5-3. Common message header for the adaptation Layer.

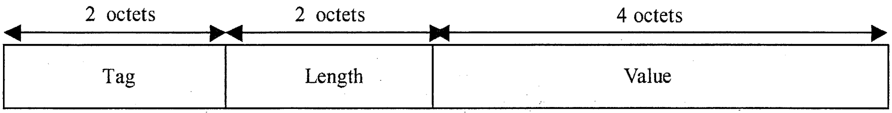


Figure 21.5-4. Variable-length parameter format for the adaptation layer.

ASP (such as a MGC or an AS). The protocol stack and the signaling architecture are shown in Fig. 21.5-5: the Q.921 protocol is terminated at the SG, and the Q.931 application runs in the ASP. IUA inserts an additional header in its messages immediately after the common header (Fig, 21.5-6). The header carries the Interface Identifier, which identifies the physical SCN interface, and the Data Link Connection Identifier (DLCI), which contains the SAPI and TEI of the LAPD interface. Q.931 messages are carried in Data messages of the QPTM class, inserted as Protocol Data parameters (Fig. 21.5-7). For a complete list of IUA messages and their formats the reader is referred to [19].

21.5.4 V5.2 User Adaptation Protocol (V5UA)

This Adaptation Layer module supports V5.2 access interfaces on the SCN side and provides backhauling of V5.2 messages between a SG and an ASP (such as a MGC or an AS). The protocol stack and the signaling architecture are shown in

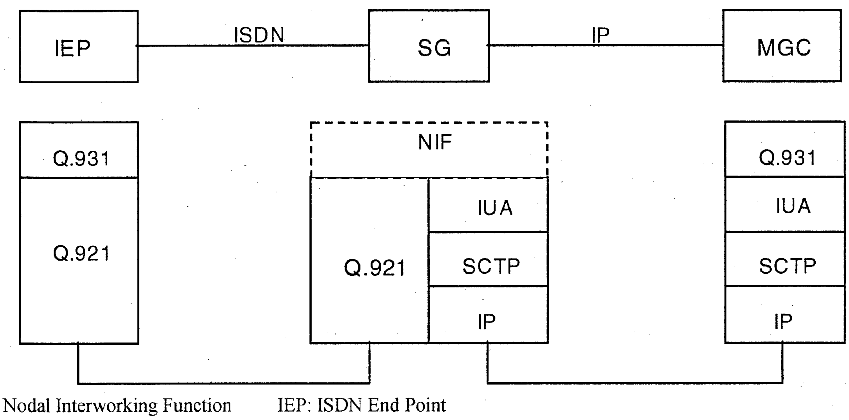


Figure 21.5-5. IUA signaling architecture. (From RFC 3057 [17].)

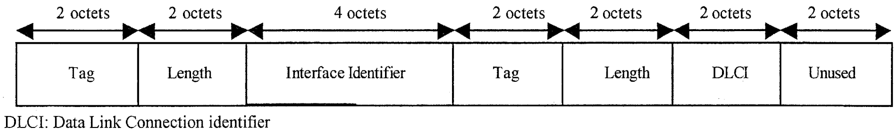


Figure 21.5-6. IUA message header.

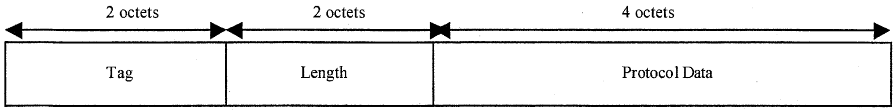


Figure 21.5-7. Protocol Data parameter for IUA Data messages.

Fig. 21.5-8: the LAPV5 protocol is terminated at the SG, and V5.2 L3 applications reside in the ASP. V5UA inserts an additional header, immediately after the common header, which is virtually the same as the IUA additional header (Fig. 21.5-6). The only difference is that the unused field is used to carry the *envelope function address* (EFA), to identify the V5.2 L3 protocols carried inside the LAPV5 PDU. The messages carrying V5.2 L3 PDUs have the same format as the corresponding messages for IUA. For a complete list of V5UA messages and their formats the reader is referred to [20].

21.5.5 SS7 MTP2 User Adaptation Layer Protocol (M2UA)

This Adaptation Layer module supports MTP2 User communication between an SS7 signaling point in the SCN and an ASP (such as a MGC or an AS) via a signaling gateway. It provides backhauling of MTP3 messages between the SG and the ASP. The protocol stack and the signaling architecture are shown in Fig. 21.5-9.

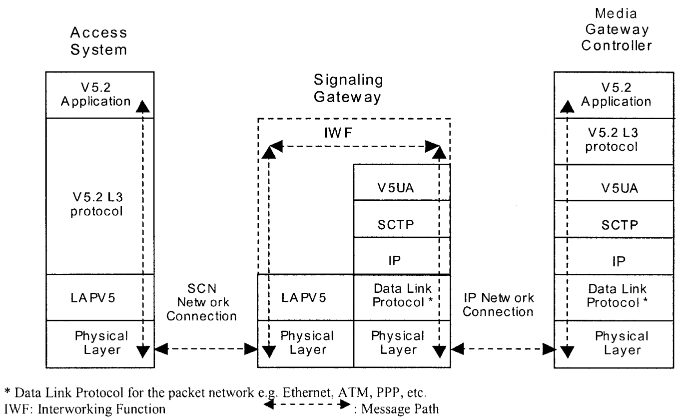
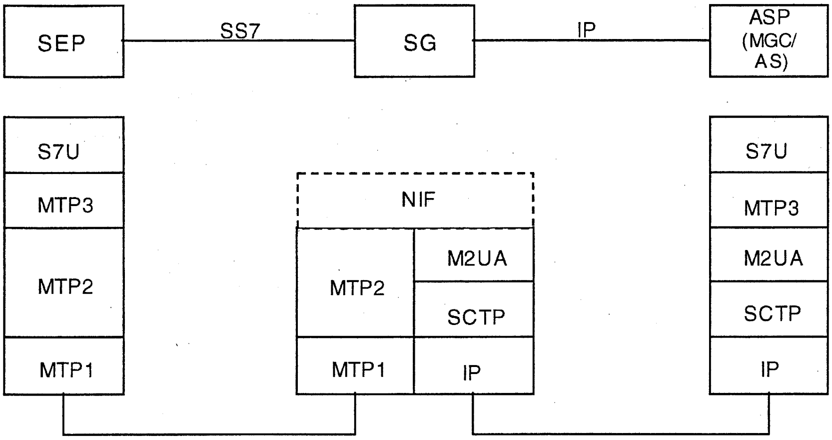


Figure 21.5-8. V5UA signaling architecture. (From RFC 3807 [18].)



NIF: Nodal Interworking Function S7U: SS7 MTP3 User

Figure 21.5-9. M2UA signaling architecture. (From RFC 3331 [19].)

MTP2 is terminated at the SG and the corresponding MTP2 User Application (MTP3) resides in the ASP. The MTP2 in the SG and MTP3 in the ASP are not aware they are running in separate nodes. The ASP has an SS7 point code; the SG does not since it contains only MTP2.

M2UA is used most effectively in network where a single central ASP is connected to multiple SGs physically remoted from one another and each serving only one SDL. In that case it is more efficient to have MTP3, and the corresponding point code, in one ASP rather than in multiple SGs.

M2UA inserts an additional header, immediately after the common header, in messages of the MTP2 User Adaptation class, to carry the Interface Identifier parameter (Fig. 21.5-10). Other message classes do not have the additional header. MTP3 messages are carried in the MAUP-class Data message, inserted as Protocol Data parameters (Fig. 21.5-11). The optional Correlation ID parameter is used to

2 octets	2 octets	4 octets
Tag	Length	Interface Identifier

Figure 21.5-10. M2UA message header.

2 octets	2 octets	4 octets	2 octets	2 octets	4 octets
Tag	Length	Protocol Data	Tag	Length	Correlation ID

Note: the Correlation ID parameter is optional

Figure 21.5-11. Parameters for the M2UA Data message.

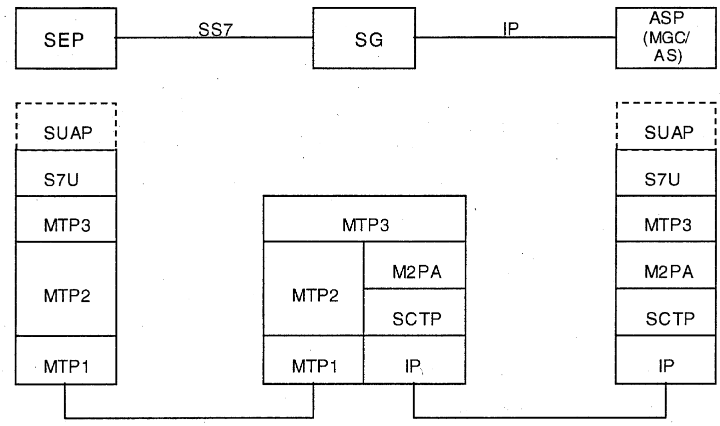
relate the carried MSU to the target ASP in case of redundant ASPs. For a complete list of M2UA messages and their formats the reader is referred to [21].

21.5.6 SS7 MTP2 User Peer-to-Peer Adaptation Layer Protocol (M2PA)

This Adaptation Layer module supports MTP2 User communication between an SS7 signaling point in the SCN and an ASP (such as MGC or AS) in the IP network, via a signaling gateway. The protocol stack and the signaling architecture are shown in Fig. 21.5-12. M2PA connections are not backhaul but peer-to-peer connections between MTP3 signaling points in the IP network. With M2PA both the SG and the ASPs connected to it have SS7 point codes and are visible to the external SS7 network. M2PA allows the smooth extension of SS7 signaling from the SCN into an IP network by using SCTP associations to emulate SDLs. M2PA actually improves on SS7 signaling data links since it supports longer sequence numbers. Network management functions for these connections use MTP3 procedures rather than Adaptation Layer (SIGTRAN) procedures.

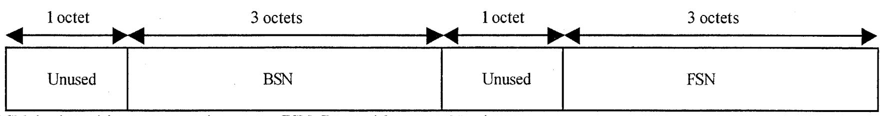
Because of its generalized approach, M2PA can be used between any type of IPSP. The SG is viewed by the SCN as a STP; therefore it may also have SCCP functionality.

M2PA inserts an additional header immediately after the common header (Fig. 21.5-13), to carry the BSN and FSN parameters of MTP2. M2PA has two



SUAP: SCCP User Application   S7U: SS7 MTP3 User

Figure 21.5-12. M2PA signaling architecture. (From RFC 4165 [20].)



BSN; backward Sequence Number   FSN: Forward Sequence Number

Figure 21.5-13. M2PA message header.

kinds of messages: User Data Messages and Link Status Messages, both of message class 11. User Data messages carry MTP3 PDUs, and Link Status messages perform functions equivalent to the LSSUs in MTP2. For a complete list of M2PA messages and their formats the reader is referred to [22].

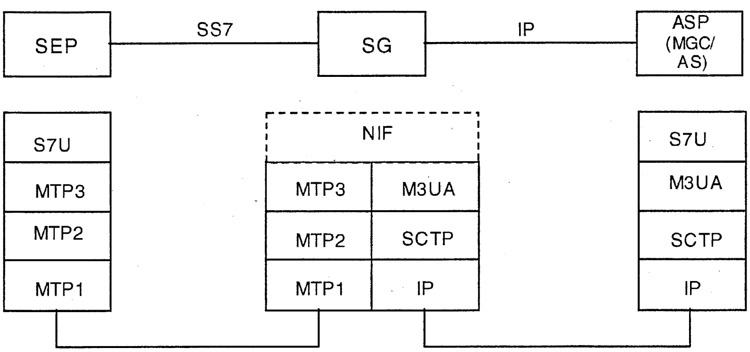
21.5.7 SS7 MTP3 User Adaptation Layer Protocol (M3UA)

This Adaptation Layer module supports MTP3 User communication between a SS7 signaling point in the SCN and an ASP (such as a MGC or an AS) via a signaling gateway. M3UA connects MTP3 to the MTP3 User by backhauling MTP3 User messages between the SG and the ASP. The MTP3 in the SG and the MTP3 User in the ASP are not aware they are hosted in separate nodes. The protocol stack and the signaling architecture are shown in Fig. 21.5-14. MTP2 and MTP3 are terminated at the SG and the corresponding MTP3 User (such as ISUP, TUP, or SCCP) runs in the ASP. The SG has an SS7 point code, since it contains MTP3, while the ASP does not.

M3UA can also be used between IPSP for connections entirely contained in the IP network, with no SG involvement.

M3UA is similar in operation to M2UA in the sense that it backhauls higher layer messages between IPSPs in client-server fashion; however, it requires MTP3 functionality and an SS7 point code in the SG. For that reason it is used most efficiently where multiple SS7 SDLs are physically terminated at one location and are connected to the same SG.

M3UA does not insert any additional header in its messages after the common header. Parameters depend on the message type. MTP3-User messages are carried in the Payload Data message (Transfer message class), which functions as a MTP TRANSFER primitive (Chapter 8). The MTP3 PDU is inserted in the message as a Protocol Data parameter (Fig. 21.5-15). For a complete list of M3UA messages and their formats the reader is referred to [23].



NIF: Nodal Interworking Function S7UP: SS7 MTP3 User

Figure 21.5-14. M3UA signaling architecture. (From RFC 3332 [21].)

4 octets	4 octets	1 octet	1 octet	1 octet	1 octet	4 octets
OPC	DPC	SI	NI	MP	SLS	UP Data

OPC: Originating Point Code DPC: Destination Point Code SI: Service Indicator NI: Network Indicator
MP: Message Priority SLS: Signaling Link Selection Code UP: User Protocol

Figure 21.5-15. Protocol Data parameter for M3UA.

21.5.8 SCCP User Adaptation Layer Protocol (SUA)

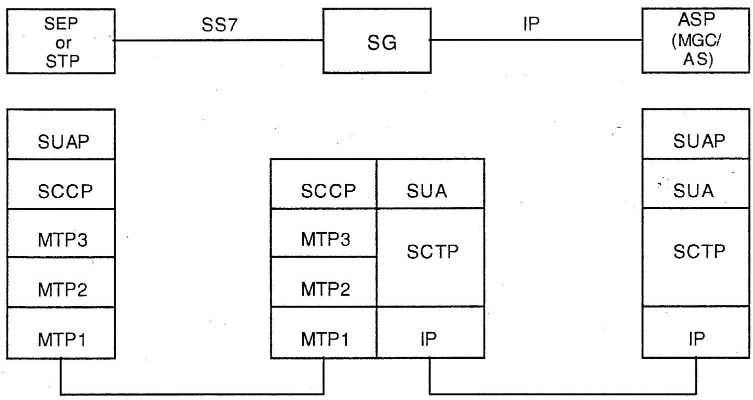
This Adaptation Layer module supports SCCP User communication between an SS7 signaling point in the SCN and an ASP (such as a MGC or AS) via a signaling gateway. The protocol stack and the signaling architecture are shown in Fig. 21.5-16. SCCP is terminated at the SG, so the SG has its own SS7 point code while the ASP does not. SUA supports only SCCP User applications such as TCAP, MAP, or RANAP, and it cannot be used for applications, like TUP and ISUP, that do not require SCCP. SUA can also be used for SCCP User communication entirely within an IP network with no SG involved.

Both connectionless and connection-oriented SCCP communication and all SCCP protocol classes are supported.

Two modes of operation are possible for a SUA node:

- Endpoint
- Relay Point

When a SG operates as an endpoint, communication between the SG and the ASP is backhaul and, from the point of view of the SS7 network, the SCCP User (e.g., TCAP) is located in the SG. The SG connects directly to the appropriate IP endpoint



SUAP: SCCP/SUA User Application

Figure 21.5-16. SUA signaling architecture. (From RFC 3868 [22].)

based on the SSN. When operating as a relay point, the SG behaves like a STP within the IP network: it may route the connection to an intermediate IP node based on SCCP and MTP3 address information and may perform GTT when receiving a GT address. Intermediate SUA nodes can also behave as relay points, resulting in multihop SCTP connections. The relay mode is optional and can be used for network scaling purposes.

For connectionless communication, each SCCP message is routed individually. For connection-oriented communication, after routing the Connection Request, the SG maps the SCCP connection on the SCN side to the SCTP association on the IP side.

SCCP User messages are transported between SUA nodes in CO and CL class messages such as *Connectionless Data Transfer* (CLDT) and *Connection-Oriented Data Transfer* (CODT). The formats for those messages are rather complex, and for their formats and a complete list of messages the reader is referred to [24].

Functionality equivalent to SUA can also be provided by M3UA, but SUA is optimized for SSCP User applications and results in a leaner protocol stack in the ASP. With SUA, only the SG requires SCCP functionality and multiple application servers are reachable from a single SG, although they do not have point codes and appear to SS7 entities in the SCN as residing in one node.

21.6 THE BEARER INDEPENDENT CALL-CONTROL (BICC) PROTOCOLS

The BICC suite of protocols is an ITU-T standard intended for control and transport of calls that transit through a packet network but are originated and terminated in the switched-circuit network (SCN).

BICC is similar in function to H.323 and SIP, but, unlike them, it does not support broadband multimedia and is designed with the key objective of providing complete service transparency for PSTN/ISDN calls transiting through the packet network. H.323 or SIP could be deployed in the same network configurations that are the target of BICC, but service transparency beyond a very basic level would not, realistically, be achievable, due to interworking difficulties.

BICC addresses the need to support traffic growth in the SCN by using packet networks for added capacity, as discussed in Section 20.3 (first avenue of convergence).

The types of packet network supported by BICC are IP (Chapter 20) and ATM (Chapter 22).

BICC specifications use the term “bearer” for user-to-user media stream, and that is the term used in the rest of this description.

21.6.1 Overview

A conceptual view of a network using the BICC protocols is shown in Fig. 21.6-1. BICC signaling and bearer transport are managed by functional entities called *service nodes* (SNs), which perform functions equivalent to those of a media

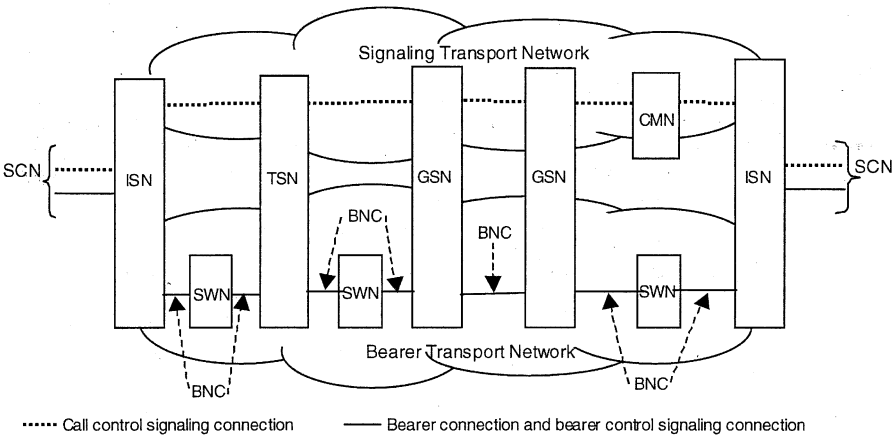


Figure 21.6-1. BICC network model.

gateway controller and a media gateway [25], thus handling both signaling and bearer channels. There are three types of service nodes:

1. Interface serving node (ISN)
2. Transit serving node (TSN)
3. Gateway serving node (GSN)

ISNs provide the direct interfaces to the SCN, TSNs are intermediate nodes in a BICC network, and GSNs connect separate BICC networks. TSNs and GSNs provide flexibility in network topology, enabling network expansion to accommodate traffic growth, and mirror the roles performed by similarly named exchange types in the SCN.

In addition to serving nodes, BICC networks also support *call mediation nodes* (CMNs), which handle only signaling channels. A CMN plays a role in IN calls, when a portion of the connection does not involve a voice path (e.g., a database query).

Another component, the *switching node* (SWN), applies only where bearers are transported by ATM. Due to the virtual-circuit nature of ATM connections, ATM switches perform a bearer-relay function—that is, they process bearer control signaling messages—and in BICC networks that function is represented by the SWN. In IP networks, the SWN is not needed because routers have no signaling role, since bearer connections are built from and known only by the originating and the terminating nodes.

The bearer connection between two adjacent service nodes is called a *backbone network connection* (BNC). A BNC may traverse SWNs.

BICC call-control signaling is based on ISUP (Chapter 11), with two differences. The first difference is that with ISUP overall call control and control of the bearer connection are integrated into the same protocol; with BICC, they are handled by separate signaling protocols and could even be transported by separate networks, as shown in

Fig. 21.6-1. That allows BICC call control to handle any type of bearer (in practice IP and ATM bearers) and is the reason for the protocol's name. The second difference is that, while ISUP signaling is transported by the SS7 signaling network (composed of signaling data links and STPs—Section 5.1), BICC call control signaling is network-independent; that is, it can be transported by an IP network or by an ATM network or even by the SS7 signaling network in the SCN. BICC bearer-control signaling, on the other hand, always uses the same transport network as the bearer connection.

21.6.2 BICC Serving Nodes

Although a serving node performs functions equivalent to those of a MGC and a MG (Section 20.3.2), its functional breakdown does not line up exactly with those two components, due to the separation of call control from bearer control. A serving node contains four functional (logical) components (Fig. 21.6-2):

1. Call serving function (CSF)
2. Bearer-control function (BCF)
3. Media-control function (MCF)
4. Media mapping and switching function (MMSF)

The CSF contains the call-control logic and communicates with peer CSFs via the BICC Call-Control Signaling Protocol. When implemented as a stand-alone physical unit, it is called a *call-control unit* (CCU).

The BCF contains the bearer-control logic and communicates with peer BCFs via a *bearer-control signaling protocol* (BCP).

The MCF is the interface between the BCF and the MMSF and allows their physical separation.

The MMSF connects bearer streams and converts dissimilar media formats. The BCF, MCF, and MMSF together form a functional entity called the *bearer interworking function* (BIWF), which groups together all bearer connection components. A CSF can control more than one BIWF.

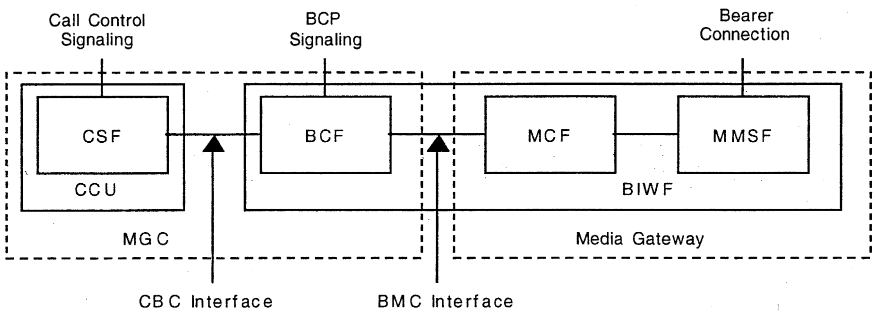


Figure 21.6-2. BICC serving node architecture.

The interface between the CSF and the BCF, called the *call bearer control* (CBC) interface, is defined and standardized in the BICC protocol suite.

The interface between the BCF and the MCF, called the *bearer and media control* (BMC) interface, is not considered part of the BICC protocol suite, and existing protocols, such as the Gateway Control Protocol/H.248 (Section 21.4), may be used for this interface.

A CMN contains only the CSF. A SWN contains a reduced-functions BCF.

Comparing the BICC serving node architecture with the VoIP network elements of Section 20.3.2, CSF and BCF correspond to the functions of a media gateway controller, while the MCF and MMSF correspond to the functions of a media gateway. The CBC interface, which is internal in a MGC, is open in the BICC architecture, allowing the BCF to be implemented as a separate physical unit from the CSF.

Implementations may merge components, such as CSF and BCF integrated into a MGC, in which case the CBC interface remains internal.

21.6.3 The BICC Protocol Suite

The BICC suite of protocols is specified in the Q.19xx series of ITU-T recommendations [26–28], which cover several protocols:

- BICC Call-Control Signaling Protocol
- BICC bearer-control signaling protocols

In order to meets the objective of PSTN/ISDN service transparency, BICC signaling was developed based on three fundamental principles:

1. Functional separation between call control and bearer control
2. Independence from the transport technology
3. No impact on existing SCN signaling interfaces

The first two principles have been discussed in the overview section; the third needs to be explained in more detail. To avoid any impact on existing SCN signaling interfaces, BICC call control signaling was derived by removing from ISUP all the functions related to bearer control and by adding *binding* functions that relate the signaling connection to the bearer connection. Because of the direct derivation from ISUP, BICC interworks seamlessly with it and does not require any modification to existing SCN exchanges.

BICC was introduced in phases, called *capability sets* (CSs):

- CS-1, which covers ATM interfaces
- CS-2, which covers IP interfaces
- CS-3, which covers interworking with SIP (not finalized at the time of this writing)

21.6.4 BICC Call-Control Signaling Protocol

The BICC Call-Control Signaling Protocol is used for peer-to-peer communication between CSFs. It follows ISUP very closely and, although not totally back-to-back compatible with it, interworking between the two is straightforward. Differences are found in the following areas:

- Message and parameter set
- Binding information
- APM encapsulation and tunnelling
- Bearer setup
- Continuity test procedures
- Codec negotiation and modification

Message and Parameter Set. The BICC message set and parameter set are virtually the same as for ISUP (Section 11.2): there are no additional message types or parameters and a few ISUP message types and parameters are not used (all those in the tables of Section 11.2 are in BICC). The slightly smaller message and parameter set is a consequence of messages and parameters related to bearer setup not being needed and their function being replaced by bearer-control protocols. One significant difference is that the CIC parameter, while retained by BICC, has a different name and function. In BICC signaling, CIC is called *Call Instance code* and identifies the instance of the signaling connection rather than the voice circuit. Also, the CIC in the BICC Call-Control Signaling Protocol is four octets long, compared to 12 bits for the ISUP CIC, thus increasing the capacity of the protocol.

Binding Information. This is used to relate a call-control instance to the corresponding bearer connection. The information elements that relate signaling to bearer are:

- CIC, assigned by the originating SN to identify the signaling instance between two nodes
- BIWF address
- BNC identity (BNC-ID), assigned by a BIWF to identify the bearer connection

APM Encapsulation. The sending and receiving CSFs exchange bearer-related information (e.g., BNC-ID and BIWF address) using the *application transport mechanism* (APM). The basic APM procedure is the same as for ISUP (Section 11.6.10) [29,30], with some BICC-specific parameter fields. The information is encapsulated in the application information field of the Application Transport parameter (Fig. 11.2-3) and included in the appropriate message, IAM or APM. The encapsulation format is shown in Fig. 21.6-3: one ATP parameter can transport several information elements, each comprised of four fields: *identifier*, which

bits → 8 7 6 5 4 3 2 1	↓octets
Identifier 1	1
Length Indicator 1	2
Compatibility Information 1	3
Contents 1	4 ...
Identifier n	m
Length Indicator n	
Compatibility Information n	
Contents n	p

Figure 21.6-3. APM encapsulation (From Rec. Q.765.5. Reproduced with the kind permission of ITU).

identifies the type of information element (e.g., BNC-ID); *length indicator*, which contains the number of octets of the IE; *compatibility information*, which indicates the action to take in case the IE is not recognized; and *contents*, which contains the actual data (e.g., the value of BNC-ID). The APM procedure is also used to "tunnel" bearer-control messages between CSFs, by encapsulating the bearer-control PDU into a BICC APM message, as described in the next section.

Bearer Setup. Two methods of bearer setup are supported by BICC:

1. Setup from the originating (preceding) SN
2. Setup from the succeeding SN

In the first case, the originating SN receives binding information from the succeeding SN in an APM message. In the second case, the succeeding SN receives binding information from the preceding SN in the IAM message. The binding information is used to start the bearer connection setup.

Continuity Test. Since bearer control is functionally separated from call control, the ISUP continuity test procedure does not apply. The Continuity Test message (COT) is still used, but it indicates to the next node that bearer setup has been completed in the preceding segment of the connection.

Codec Negotiation and Modification. This is an optional enhancement over ISUP, to support wireless communication. A prioritized list of codecs is sent by the originating ISN in the IAM message. The terminating ISN picks one codec out of the list and sends the choice back in an APM message. Codec negotiation works end-to-end and, while being performed, bearer setup (done link-by-link) is suspended.

21.6.5 Bearer-Control Signaling Protocols

Bearer-control protocols are used by BCFs to exchange information, such as bearer characteristics, port number, and IP address of the originating and terminating user interface, necessary to set up bearer connections. BICC recommendations cover several signaling protocols for bearer control, in support of two functions:

1. Intranode control of the BCF by the CSF, through the CBC interface
2. Internode (peer-to-peer) communication between BCFs, to set up and release bearer connections

Intranode Protocol. One signaling protocol is specified for the CBC interface: the Bearer Independent Call Bearer-Control Protocol. It is essentially the same as the H.248 protocol (Section 21.4), with minor extensions not discussed here. For the details we refer the reader to [16,31].

Internode Protocols. Internode bearer-control signaling protocols, generically called bearer-control protocols (BCPs), are bearer-specific, so different ones are used for IP and ATM networks.

IP Networks. The bearer-control protocol specified by ITU-T for IP networks is the IP Bearer-Control Protocol (IPBCP) [32]. It is very similar to SDP (Section 20.4.3) [11], with some IPBCP-specific parameters added. Four message types are used, identified by the value of the <type> field in the *Session Attribute* parameter:

1. Request
2. Accepted
3. Confused
4. Rejected

Parameters in an IPBCP message convey bearer connection information, such as network (IP) address, port number, and media type.

The protocol operation is straightforward: a Request message is sent by the originating BIWF, containing its media interface address. The receiving BIWF responds with an Accepted message containing its media interface address, at which point the media connection is established. Media channels established by IPBCP are bidirectional.

The IPBCP recommendation leaves the choice of transport-layer protocols open to any reliable, point-to-point communication arrangement. One option that is explicitly allowed is for messages to be exchanged between BCFs by "tunneling" through the CBC (if CSF and BMF are physically separate) and the BICC call-control interfaces. The procedure is called *tunneling* because IPBCP messages pass through the BMF–CSF and CSF–CSF interfaces as payload encapsulated in messages of other protocols.

bits →	8	7	6	5	4	3	2	1	↓octets
	0	BVEI	1	BCTP Version Indicator					1
	0	TPEI	Tunneled Protocol Indicator						2

BVEI: BCTP Version Error Indicator TPEI: Tunneled Protocol Error Indicator

Figure 21.6-4. BCTP encapsulation header (From Rec. Q.1990. R reproduced with the kind permission of ITU).

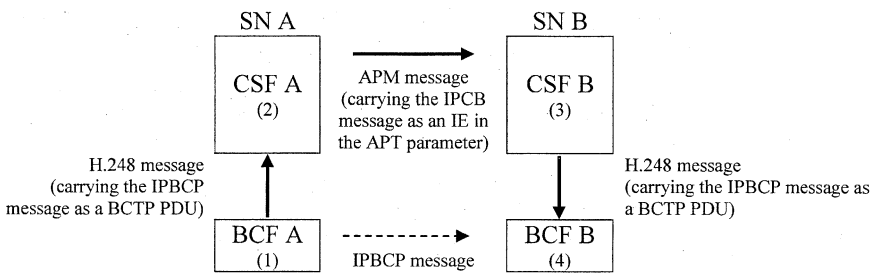
Messages are first encapsulated by the sending BCF using the Bearer-Control Tunneling Protocol (BCTP) [33], a generic tunneling procedure for transporting bearer-control messages. Encapsulation inserts a header with the format shown in Fig. 21.6-4. The resulting BCTP PDU is then sent from the BCF to the CSF using the H.248 tunneling mechanism (not discussed here [16]), and from the CSF to the peer CSF using the APM procedure (see previous section). At the destination CSF, the process is reversed. The tunneling process is shown in Fig. 21.6-5.

ATM Networks. ATM networks use ATM protocols for bearer channel setup, such as UNI, PNNI, and B-ISUP, discussed in Chapter 22. Tunneling is not used.

21.6.6 BICC Call Setup

BICC call setup involves the interaction of all entities in the serving node architecture of Fig. 21.6-2 and the interworking of all BICC protocols.

An example of call setup is shown in Fig. 21.6-6. In the example bearer channels are set up from the originating (preceding) SN, and all SN components communicate via protocol messages, except for the MCF–MMSF interface, assumed to be internal. The example shows a call setup sequence between two adjacent SNs,



→ Physical message path ----→ Logical message path

(1) IPBCP message is encapsulated in a BCTP PDU and inserted into a H.248 message.
(2) IPBCP message is extracted from the H.248 message, encapsulated in an APT parameter and inserted into a BICC APM message.
(3) IPBCP message is extracted from the APM message, encapsulated in a BCTP PDU and inserted into a H.248 message.
(4) IPBCP message is extracted from the H.248 message.

Figure 21.6-5. IPBCP tunneling.

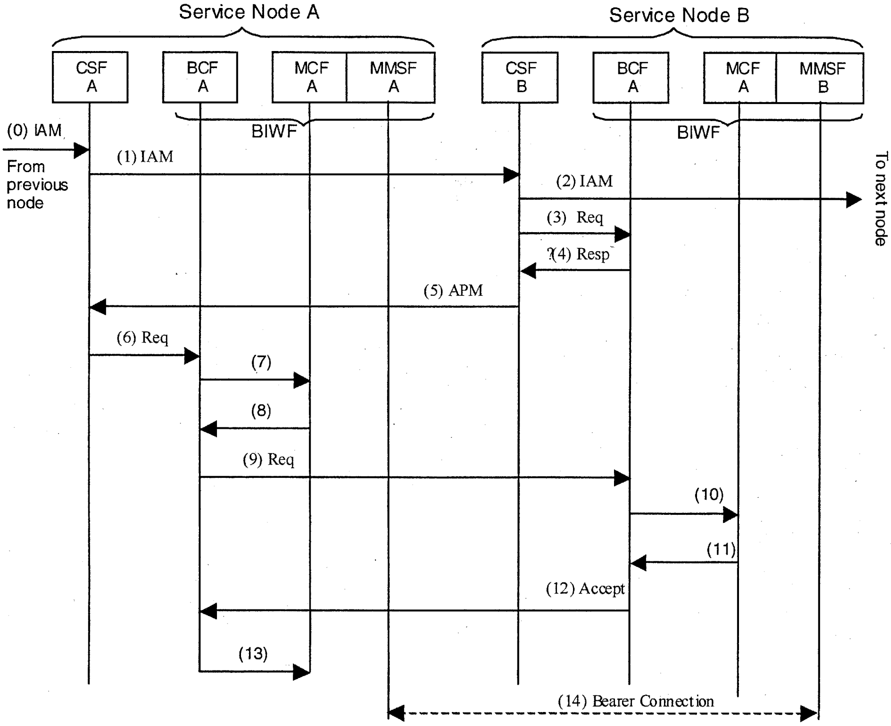


Figure 21.6-6. BICC call setup example.

where SN-A is the originating (preceding) SN and SN-B is the terminating (succeeding) SN. The example describes a link in an end-to-end connection and applies equally to ISN–TSN, TSN–TSN, TSN–GSN, or GSN–GSN connections.

The process starts (0) when the CSF in SN-A (CSF-A) receives an IAM message, which could be an ISUP IAM (for an ISN), or a BICC IAM (for other SN types). CSF-A selects a BIWF and then processes the IAM to determine the routing of the call. Once the route has been chosen (in this case CSF-B), CSF-A sends (1) a BICC IAM message to CSF-B with the media characteristics of its BIWF. CSF-B processes the IAM and, if able to route the call, sends (2) an IAM to the next node. CSF-B then selects a BIWF that matches the media characteristics from the received IAM and (3) requests, and (4) receives from BCF-B (via the CBC protocol), the BCN-ID and the media characteristics of the bearer connection. CSF-B sends (5) BIWF-B's address, the BNC-ID, and the media characteristics to CSF-A in an APM message. CSF-A (6) forwards the binding information to BCF-A in a CBC protocol message. BCF-A then (7) requests, and (8) receives from MMSF-A, the bearer interface address, via a BMC protocol. BCF-A (9) sends MMSF-A's bearer address to BCF-B, who forwards it (10) to MMSF-B. MMSF-B sends (11) its own bearer address to BCF-B, from where (12) it is forwarded to BCF-A. BCF-A (13) forwards MMSF B's bearer address to MMSF-A.

At this point MMSF-A and MMSF-B have each other's bearer address and can (14) establish voice communication. After the media connection is established, BIWF-A informs CSF-A and BIWF-B informs CSF-B of the completion of the bearer connection. Optionally, CSF-A and CSF-B can send confirmation to one another via an APM message (not shown). If the sending of the COT message is enabled, CSF-B issues a COT message to the next node. The connection is built link-by-link and when the terminating exchange is connected, Address Complete (ACM) and Answer (ANM) messages are sent back through the nodes as described in Chapter 11 for ISUP.

21.7 ACRONYMS

ABNF	Augmented Bakus–Naur form
APM	Application transport mechanism
AS	Application server
ASN.1	Abstract Syntax Notation One
ASP	Application server process
ATP	Application transport parameter
BCP	Bearer-control protocol
BCF	Bearer-control function
BCTP	Bearer-Control Tunneling Protocol
BE	Border element
BER	Basic encoding rules
BICC	Bearer Independent Call Control
BIWF	Bearer interworking function
BMC	Bearer and media control
BNC	Backbone network connection
BNC-ID	Backbone network connection identifier
BSN	Backward sequence number
CAS	Channel-associated signaling
CBC	Call bearer control
CCU	Call-control unit
CIC	Call instance code
CLDT	Connectionless data transfer
CMN	Call mediation node
CODT	Connection-oriented data transfer
COT	Continuity test
CRLF	Carrier return followed by line feed
CS	Capability set
CSF	Call serving function
CSRC	Contributing sources
DHCP	Dynamic Host Configuration Protocol
DLCI	Data link connection identifier
DNS	Domain name system
EFA	Envelope function address

FSN	Forward sequence number
GK	Gatekeeper
GSN	Gateway serving node
GT	Global title
GTT	Global title translation
GW	Gateway
HTTP	Hypertext Transfer Protocol
IAM	Initial Address message
ID	Identity
IETF	Internet Engineering Task Force
INAP	Intelligent Network Application Part
IP	Internet Protocol
IPBCP	IP Bearer-Control Protocol
IPSP	IP signaling point
ISDN	Integrated Services Digital Network
ISN	Interface serving node
ISUP	ISDN User Part
ITU	International Telecommunications Union
ITU-T	Telecommunication Standardization Sector of ITU
IUA	ISDN user adaptation
L3	Layer 3
L4	Layer 4
LAPV5	Link access protocol V5
M2PA	Message transfer part 2 user peer-to-peer adaptation
M2UA	Message transfer part 2 user adaptation
M3UA	Message transfer part 3 user adaptation
MAP	Mobile Application Part
MCF	Media-control function
MCU	Multipoint control unit
MG	Media gateway
MGC	Media gateway controller
MGCP	MGC protocol
MID	Message identity
MMSF	Media mapping and switching function
MTP	Message Transfer Part
MTP2	MTP Level 2
MTP3	MTP Level 3
MTU	Maximum transmission unit
NGN	Next-generation network
NIF	Nodal interworking function
OSI	Open Systems Interconnection
PC	Personal computer
PDU	Protocol data unit
QoS	Quality of service
RANAP	Radio Access Network Application Part

RAS	Registration, admission, and status
RFC	Request for comment
RTCP	RTP Control Protocol
RTP	Real-Time Transport Protocol
SCCP	Signaling Connection Control Part
SCN	Switched-circuit network
SCTP	Stream Control Transmission Protocol
SDL	Signaling data link
SDP	Session Description Protocol
SET	Simple endpoint type
SG	Signaling gateway
SIGTRAN	Signaling Translation
SIP	Session Initiation Protocol
SN	Service node
SP	Space (character)
SS7	Signaling System No. 7
STP	Signal transfer point
SUA	SCCP user adaptation
SWN	Switching node
TCAP	Transaction Capabilities Application Part
TCP	Transmission Control Protocol
TDM	Time-division multiplexing
TSN	Transit serving node
TUP	Telephone User Part
UA	User agent
UAC	User agent client
UAS	User agent server
UDP	Used Datagram Protocol
ULP	Upper layer protocol
URI	Uniform resource identifier
URL	Uniform resource locator
V5UA	V5.2 user adaptation
VoIP	Voice over IP

21.8 REFERENCES

1. *Packet-Based Multimedia Communications Systems*, Rec. H.323, ITU-T, Geneva, July 2003.
2. *Call Signalling Protocols and Media Stream Packetization for Packet-Based Multimedia Communication Systems*, Rec. H.225, ITU-T, Geneva, July 2003.
3. *Control Protocol for Multimedia Communication*, Rec. H.245, ITU-T, Geneva, July 2003.
4. *Security and Encryption for H-Series (H.323 and Other H.245-Based) Multimedia Terminals* Rec. H.235, ITU-T, Geneva, Aug. 2003.
5. *Information Technology—Abstract Syntax Notation One (ASN.1): Specification of Basic Notation*, Rec. X.680, ITU-T, Geneva, July 2002.

6. *Generic Functional Protocol for the Support of Supplementary Services in H.323*, Rec. H.450.1, ITU-T, Geneva, Feb. 1998.
7. *Hypertext Transfer Protocol—HTTP/1.1*, RFC 2068, IETF, Jan. 1997.
8. *SIP: Session Initiation Protocol*, RFC 3261, IETF, June 2002. Copyright (c) The Internet Society (2002). All rights reserved. This document and translations of it may be copied and furnished to others, and derivative works that comment on or otherwise explain it or assist in its implementation may be prepared, copied, published and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice and this paragraph are included on all such copies and derivative works. However, this document itself may not be modified in any way, such as by removing the copyright notice or references to the Internet Society or other Internet organizations, except as needed for the purpose of developing Internet standards in which case the procedures for copyrights defined in the Internet Standards process must be followed, or as required to translate it into languages other than English. The limited permissions granted above are perpetual and will not be revoked by the Internet Society or its successors or assigns.
9. *Session Initiation Protocol (SIP) Basic Call Flow Example*, RFC 3665, IETF, Dec. 2003.
10. *Session Initiation Protocol (SIP): Locating SIP Servers*, RFC 3263, IETF, June 2002.
11. *SDP: Session Description Protocol*, RFC 2327, IETF, Apr. 1998.
12. *Augmented BNF for Syntax Specifications: ABNF*, RFC 2234, IETF, Nov. 1997.
13. *Uniform Resource Identifiers (URI): Generic Syntax*, RFC 2396, IETF, Aug. 1998.
14. *URLs for Telephone Calls,* RFC 2806, IETF, Apr. 2000.
15. *Gateway Control Protocol Version 1*, RFC 3525, IETF, June 2003. Copyright (c) The Internet Society (2003). All rights reserved. (The rest of the copyright statement is the same as for [8].)
16. H.248-series recommendations (H.248.1 through H.248.35), ITU-T, Geneva, 2000–2005.
17. *Media Gateway Control Protocol (MGCP)*, RFC 2705, IETF, Oct. 1999.
18. *Framework Architecture for Signaling Transport*, RFC 2719, IETF, Oct. 1999.
19. *ISDN Q.921-User Adaptation Layer*, RFC 3057, IETF, Feb. 2001. Copyright (c) The Internet Society (2001). All rights reserved. (The rest of the copyright statement is the same as for [8].)
20. *V5.2-User Adaptation Layer (V5UA)*, RFC 3807, IETF, June 2004.
21. *Signaling System 7 (SS7) Message Transfer Part 2 (MTP2)—User Adaptation Layer*, RFC 3331, IETF, Sept. 2002. Copyright (c) The Internet Society (2002). All rights reserved. (The rest of the copyright statement is the same as for [8].)
22. *Signaling System 7 (SS7) Message Transfer Part 2 (MTP2)—User Peer-to-Peer Adaptation Layer (M2PA)*, RFC 4165, Sept. 2005.
23. *Signaling System 7 (SS7) Message Transfer Part 3 (MTP3)—User Adaptation Layer (M3UA)*, RFC 3332, IETF, Sept. 2002. Copyright (c) The Internet Society (2002). All rights reserved. (The rest of the copyright statement is the same as for [8].)
24. *Signalling Connection Control Part User Adaptation Layer (SUA)*, RFC 3868, IETF, Oct. 2004.
25. R. R. Knight, S. E. Norreys, and J. R. Harrison, Bearer-Independent Call Control, *BT Technol. J.* **19** No. 2, Apr. 2001.
26. *Bearer Independent Call Control Protocol,* Rec. Q.1901, ITU-T, Geneva, June 2000.

27. *Bearer Independent Call Control Protocol (Capability Set 2): Functional Description*, Rec. Q.1902.1, ITU-T, Geneva, July 2001.
28. *Bearer Independent Call Control Protocol (Capability Set 2) and Signalling System No. 7 ISDN User Part: General Functions of Messages and Parameters*, Rec. Q.1902.2, ITU-T, Geneva, July 2001.
29. *Signalling System No. 7—Application Transport Mechanism: Bearer Independent Call Control (BICC)*, Rec. Q.765.5, ITU-T, Geneva, Apr. 2004.
30. *Signalling System No. 7—Application Transport Mechanism*, Rec. Q.765, ITU-T, Geneva, June 2000.
31. *Bearer Independent Call Bearer Control Protocol*, Rec. Q.1950, ITU-T, Geneva, Dec. 2002.
32. *BICC IP Bearer Control Protocol*, Rec. Q.1970, ITU-T, Geneva, July 2001.
33. *BICC Bearer Control Tunnelling Protocol*, Rec. Q.1990, ITU-T, Geneva, July 2001.

22

SIGNALING IN ATM NETWORKS

ATM, or *asynchronous transfer mode*, is a digital technology for voice and data convergence based on short, fixed-length packets called *cells* [1,2]. For that reason ATM is sometimes called *cell relay*. ATM had its start in the international standards work for *broadband ISDN* (B-ISDN), sponsored by ITU-T in the 1980s.

Two standards bodies (see Section 2.2) issue ATM specifications:

1. ITU-T, which issues B-ISDN recommendations, focusing primarily on public service providers and telecommunication administrations.
2. MFA (MPLS, Frame Relay, ATM) Forum, which issues ATM specifications focusing primarily on private networks. The MFA Forum specifications mentioned in this chapter were developed by the ATM Forum prior to its 2005 merger with MFA, and we refer to them as ATM Forum (AF) documents.

AF specs are based on ITU-T recommendations and are often specified as deltas to the corresponding ITU-T documents; in some cases they also borrow from the IETF (Internet Engineering Task Force).

Since ATM is a form of packet technology, many of the considerations discussed in Chapter 20 apply to it. We refer the reader to that chapter for the basic concepts of packet communication, such as OSI layers, routers and switches, connection-oriented versus connectionless, jitter, delays, and queues.

22.1 INTRODUCTION TO ATM NETWORKS AND INTERFACES

An ATM network is a collection of nodes, called *switches*, interconnected by digital links and providing communication between *ATM endpoints* (Fig. 22.1-1). An ATM endpoint (sometimes called an *edge device*) can be an ATM user terminal or a

Signaling in Telecommunication Networks, Second Edition, by John G. van Bosse and Fabrizio U. Devetak

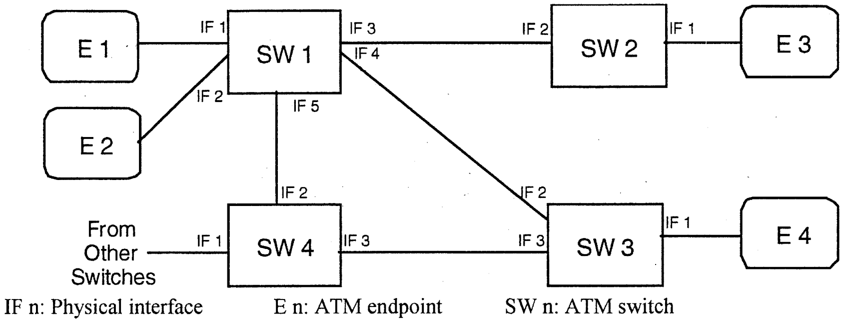


Figure 22.1-1. Example of ATM network configuration.

gateway between the ATM network and another type of network. In an ATM network there are two types of interfaces:

1. User–network interface (UNI), between an ATM endpoint and an ATM switch. This interface is also called the *access interface*.
2. Network–network or network–node interface (NNI), between two ATM switches. This interface is also called the *network interface*.

An ATM network can be owned by a telecommunication company (carrier), in which case it is called a *public ATM network*; or it can be owned by an enterprise, in which case it is called a *private ATM network*.

Depending on the entities owning the endpoints and the network nodes, NNI and UNI interfaces can be private or public. Private interfaces are between entities in a private network; public interfaces are between entities in a public network, or between a private and a public network, or between two public networks (Fig. 22.1-2).

ATM was originally intended as an alternative to IP and Ethernet, to be deployed seamlessly from public backbone networks all the way to LANs and residential users [2]. However, after losing the desktop war to Ethernet and IP, ATM is now deployed mainly on xDSL lines (Section 2.1.5) and in carrier backbone networks that interconnect IP networks.

The transfer mode used by ATM is called asynchronous because the recurrence of cells in a data flow is not necessarily periodic, and cells do not have a preassigned temporal position or time slot in a transmission link as in time-division multiplexing (TDM—Section 1.5). In an ATM cell, a header identifying a particular data stream is prepended to the user data (payload) and allows each cell to be routed individually regardless of its position in the bit stream, thus enabling *statistical multiplexing* of data flows in a transmission link.

ATM achieves two main objectives in voice and data convergence:

- Broadband multimedia communication
- Easy interworking between packet and circuit-switched networks

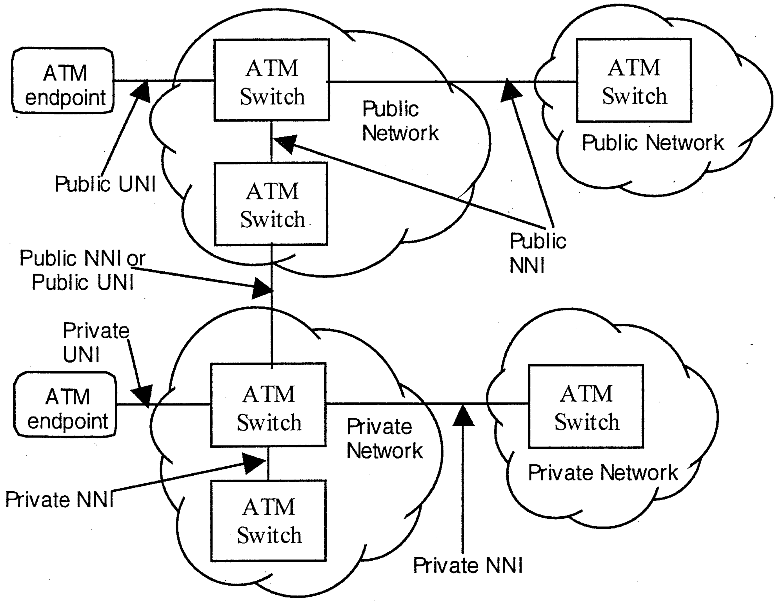


Figure 22.1-2. UNI and NNI interfaces.

Cell characteristics ensure superior latency and jitter characteristics (compared to IP-based packet communication), making ATM well suited not only for data but also for real-time multimedia applications such as voice and video. The same characteristics make highly efficient switching of broadband data flows possible: ATM systems can easily support rates of up to multigigabits per second [3].

ATM is connection-oriented: a virtual (logical) connection with a fixed path is established by signaling messages at the start of a data flow and released at the end, thus ensuring cell sequence integrity. Thanks to virtual connections and the predictability of fixed-length cells, ATM can emulate traditional TDM circuits when configured to transfer cells at a fixed rate. Circuit emulation introduces a small performance penalty, compared to TDM, due to header overhead, but is of interest to service providers with legacy TDM circuits who want to phase-in packet technology and minimize interworking problems.

As in TDM, no checks are done on the payload, which is transferred transparently from source to destination.

ATM supports variable data rates by allocating bandwidth based on demand and transmitting cells only when there is data to send. Cells can be sent on unframed transmission systems in cell-based mode, where cells are placed on the data link back-to-back; or they can be sent on framed systems where a frame structure is present. Examples of framed systems are the Synchronous Digital Hierarchy (SDH) and the Synchronous Optical Network (SONET), discussed in Section 1.5.3.

ATM also supports elaborate service class and bandwidth management capabilities, which ensure efficient handling of data rates with both predictable and unpredictable bandwidth characteristics, as well as strict enforcement of *quality of service* (QoS) [4,5].

In summary, ATM incorporates the best aspects of packet and circuit-switched technologies, providing a generalized transport and networking framework that can handle any kind of data source. The higher complexity of the requirements and the higher cost of ATM equipment, however, have shifted the momentum for voice and data convergence in the direction of TCP/IP (Chapter 20).

22.1.1 Virtual Channel and Virtual Path Connections

ATM communication is based on two types of virtual connections:

- Virtual channel connection (VCC)
- Virtual path connection (VPC)

Virtual channel connection is the ATM term for what is generally called a *virtual circuit*, that is, a logical connection between two nodes or endpoints that agree to exchange packets based on parameters negotiated during connection setup. A virtual circuit is a form of connection-oriented packet communication (Section 20.1.3) in which the communication path stays fixed for the duration of the connection, thus ensuring packet sequence integrity.

A VCC can traverse multiple intermediate nodes between the origination and the termination. A segment of a virtual channel connection between two adjacent nodes is called a *virtual channel link* (VCL). VCLs are identified by a parameter in the cell header called the *virtual channel identifier* (VCI). VCIs have only local significance: the same VCC may have a different VCIs in each VCL. The establishment of a VCC requires the preexistence a VPC (below).

A virtual path connection is a collection of VCCs (or of VCC segments) that are switched and managed as a group. A VPC can traverse multiple nodes between the origination and the termination. A segment of a virtual path connection between two adjacent nodes is called a *virtual path link* (VPL). VPLs are identified by a parameter in the cell header called the *virtual path identifier* (VPI). VPIs have only local significance: the same VPC may have a different VPI in each VPL. Virtual path connections provide several benefits:

- Efficient routing
- Management of traffic and communication resources for QoS
- Fast recovery in case of path failure

Efficient routing results from the fact that intermediate nodes need not be involved in call setup when a VCC is assigned to a preexisting VPC.

Traffic capacity and communication resources can be reserved for VPCs, so they can be used to consolidate and manage traffic with similar characteristics.

VPCs also allow fast recovery in case of path failure, since an alternate path can be brought up almost instantly without interrupting the data flow, as there is no need to cycle through all the VCCs individually.

In traditional PSTN terminology, VCLs are similar in function to trunks, VCCs to end-to-end connections, VPLs to trunk groups, and VPCs to routes.

VCI values are unique only within a VPL, and VPI values are unique only within a physical interface. A VCC is uniquely identified, on a given physical link, by the paired values of VPI and VCI.

Both VCLs and VPLs, and hence VCCs and VPCs, are unidirectional; ATM communication, however, is bidirectional, composed of paired unidirectional VCCs and VPCs (one per direction), both following the same path through the network. Since transmission parameters can be set up independently for each direction, a unidirectional connection can be obtained simply by setting the bandwidth in one direction to zero. One exception is ATM's support of point-to-multipoint connections. These are unidirectional connections from one source (the *root*) to multiple destinations (the *leaves* in a *tree*). ATM cells are issued once at the root and then replicated as necessary by the network for delivery to the leaves [6]. Possible applications are video or audio distribution.

VCCs and VPCs can be established semipermanently by provisioning or dynamically via ATM signaling. In the first case, they are called *permanent virtual channel connections* (PVCCs) and *permanent virtual path connections* (PVPCs); in the second case, they are called *switched virtual channel connections* (SVCCs) and *switched virtual path connections* (SVPCs). VPCs are most commonly setup semipermanently by network administrators.

VCCs and VPCs can be established for user-to-user, user-to-network, or network-to-network connections.

22.1.2 ATM Switching

There are two types of ATM switches: *virtual path switches*, which switch VPCs based on VPI values with VCI values left unchanged, and *virtual circuit switches*, which switch VCCs based on both VCI and VPI values. Virtual path switches are also called *ATM cross-connects*; VCI switches are also called *broadband switching systems* (BSSs). Figure 22.1-3 illustrates the operation of the two types of switches with reference to the network configuration of Fig. 22.1-1: (a) is an example of virtual path switching inside SW 4 and (b) is an example of virtual circuit switching inside SW 1.

ATM switching is different from packet switching as done by IP routers and consists of a three-stage process:

1. Establish a virtual connection, assigning VCI and/or VPI labels for each link. This is accomplished by signaling procedures (Section 22.3).
2. Switch payload-carrying cells based on VCI and VPI labels.
3. Release the virtual connection at the end of the communication process. This is also accomplished by signaling procedures.

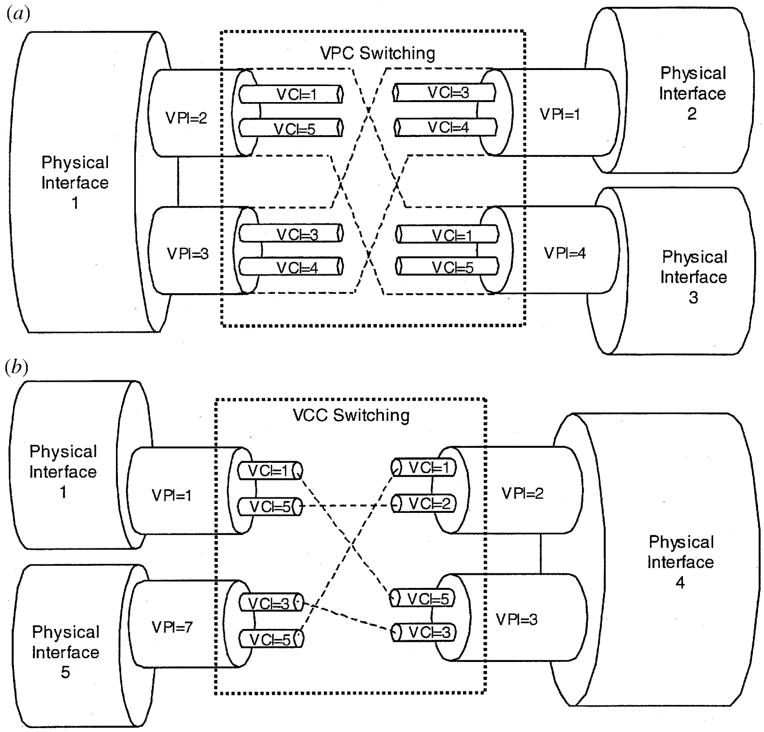


Figure 22.1-3. Virtual Path Connection and Virtual Channel Connection switching.

ATM uses global logical addresses, functionally similar to IP addresses (Section 22.9) for the first stage (connection setup), and then uses the combination of VPI and VCI labels during user-data communication, to route cells link-by-link. In traditional packet switching, a global logical address is used to route every packet at every hop toward the destination, in what is called the *hop-by-hop paradigm.* The use of logical addressing for routing each individual packet in IP networks has a heavier impact on processing resources because the address pool is much larger and the address itself is typically much longer than labels with only local significance.

VPI and VCI values are assigned locally (link-by-link) by ATM switches during connection setup and, for a given connection, the same pair of values is assigned for each direction and at each side of a link.

The paired values of VCI and VPI, which identify a VCC on a physical link and which are carried in the header of each cell, enable a switch to route an incoming cell very efficiently using a *translation table*. The translation table is populated by a switch during call setup; the switch then uses the table to forward each cell from the incoming to the outgoing link and to translate the incoming VPI/VCI values into the outgoing VPI/VCI values. The VPI and VCI values in the header of the received cell are replaced, before sending, by the values for the outgoing link.

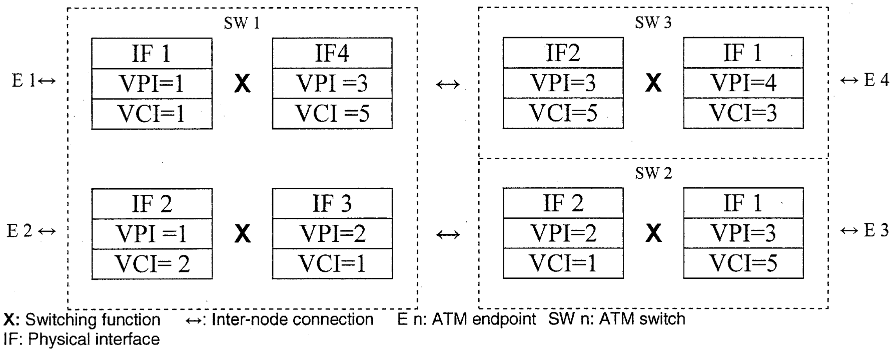


Figure 22.1-4. Example of routing tables for ATM cell switching.

A conceptual example of cell switching, based on the network configuration of Fig. 22.1-1, is shown in Fig. 22.1-4. In Fig. 22.1-4, E 1 is connected to E 4, and E 2 to E 3. SW 1 receives cells from E 1 on physical interface IF 1 and forwards them to physical interface IF 4, changing their VPI/VCI values from 1/1 to 3/5. SW 3 receives cells on its IF 2 physical interface and forwards them to its IF 1 physical interface, relabeling them from 3/5 to 4/3. Operation in the reverse direction takes place in reverse order. The E 2–E 3 connection works the same way, with its own VCI and VPI values.

22.2 ATM LAYERS AND PROTOCOL STACK

ITU-T standards define an ATM layered reference model (Fig. 22.2-1 [7]), structured along three vertical *planes* and four horizontal *layers*.

The three planes are:

- The *user plane*, which supports transfer of user (bearer) information.
- The *control plane*, which supports connection control (signaling).
- The *management plane* is further subdivided into *layer management*, which supports management of layer resources, and *plane management*, which supports coordination between planes. The management plane is not discussed in this chapter [4,5].

Planes use the services of, or provide services to, all layers. In virtual circuit and virtual path connections, communication between source and destination is first established in the control plane. After acceptance by the destination, communication takes place in the user plane.

The four layers are conceptually similar to the layers of the OSI model (Section 20.1.3). The *physical layer* supports transmission media, bit encoding, and cell delineation, corresponding to L1 of the OSI model. The *ATM layer* manages cell

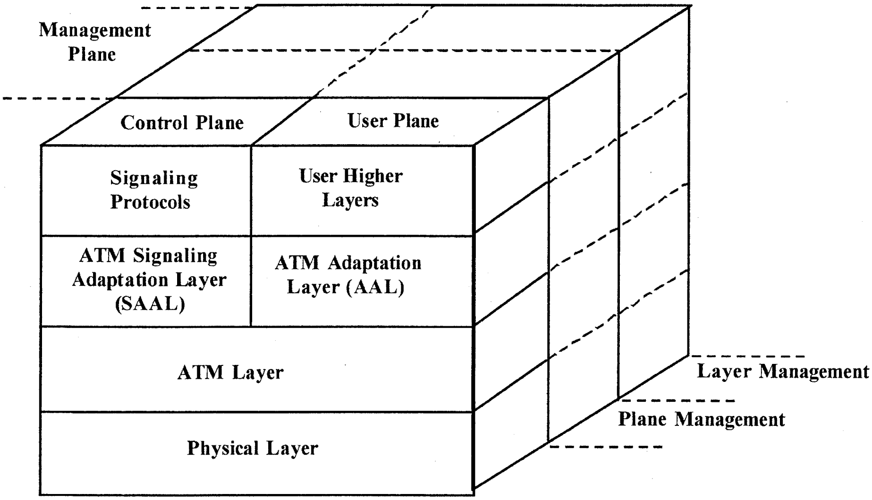


Figure 22.2-1. ATM reference model.

headers and supports cell transfer (e.g., switching and multiplexing), corresponding to L2 of the OSI model. The *ATM adaptation layer* (AAL) breaks up the data flow from the source (e.g., packets, or circuit-switched data) into 48-byte payload units that fit into cells and reassembles it at the destination. It corresponds to L3 and L4 of the OSI model. The AAL in the control plane is slightly different from the AAL in the user plane and is called *signaling ATM adaptation layer* (SAAL) [8,9]. The higher layers support call control and traffic management and correspond to L7 of the OSI model.

In a mixed IP and ATM environment, when looking at the ATM protocol from the perspective of an IP network, ATM appears as a Layer 2 protocol, since an ATM network provides the equivalent of point-to-point data links (e.g., WAN links) to the IP network. ATM network internals—how an IP packet is transported from one end of an ATM VCC to the other—are invisible to the IP network [10].

AAL and SAAL are subdivided into the same sublayers, shown in Fig. 22.2-2: the *convergence sublayer* (CS) and the *segmentation and reassembly* (SAR) *sublayer*.

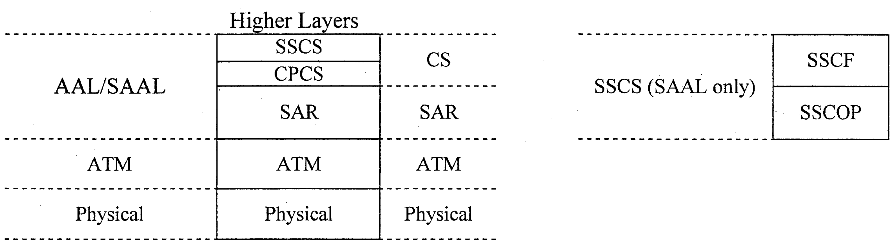


Figure 22.2-2. AAL/SAAL sublayers.

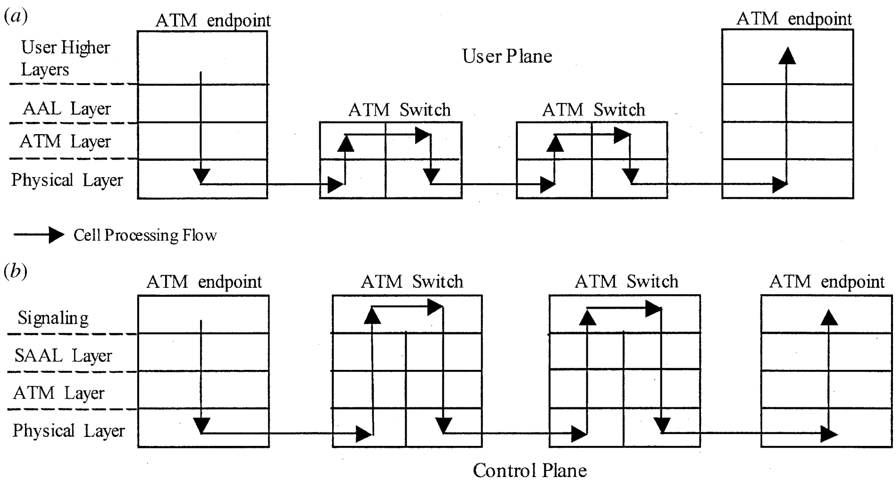


Figure 22.2-3. ATM layers and ATM nodes in the User Plane and in the Control Plane.

The CS has to adapt to the characteristics of the sources, so is itself subdivided into a *service-specific CS* (SSCS) and a *common part CS* (CPCS). The only difference between AAL and SAAL is that, for SAAL, the SSCS is further subdivided into two sublayers: the Service-Specific Connection-Oriented Protocol (SSCOP) [11] and the Service-Specific Coordination Function (SSCF) [12,13].

Figure 22.2-3(*a*) shows the layers of the ATM reference model involved in VCC payload transfer (bearer traffic in the user plane) in the different types of ATM entities. For signaling (i.e., in the control plane, during connection setup or release), endpoints and intermediate nodes are involved at all layers of the mode (Fig. 22.2-3(*b*)), for reasons that will become clear in the next sections.

22.3 LOWER LAYERS

22.3.1 ATM Layer

The ATM layer is responsible for cell formatting. An ATM cell consists of a 5-byte header and a 48-byte information field, for a total of 53 bytes. The size of the information field is the result of a compromise reached at the then CCITT between European and North American proposals. Both sides wanted a short cell for speed and efficiency, but Europeans advocated a 32-byte payload, while North Americans favored a 64-byte payload; so it was agreed to split the difference. The header contains the information necessary for switching cells and the information field contains the data payload, which is transferred transparently from source to destination. The cell headers for the UNI and the NNI interfaces are slightly different. Figures 22.3-1 and 22.3-2 show the ATM cell formats for each type of interface.

<table>
<tr><th></th><th>Bits→
Bytes↓</th><th>8</th><th>7</th><th>6</th><th>5</th><th>4</th><th>3</th><th>2</th><th>1</th></tr>
<tr><td rowspan="5">Header</td><td>1</td><td colspan="4">GFC</td><td colspan="4">VPI</td></tr>
<tr><td>2</td><td colspan="4">VPI</td><td colspan="4">VCI</td></tr>
<tr><td>3</td><td colspan="8">VCI</td></tr>
<tr><td>4</td><td colspan="4">VCI</td><td colspan="3">PT</td><td>CLP</td></tr>
<tr><td>5</td><td colspan="8">HEC</td></tr>
<tr><td>Payload</td><td>6
7
...
53</td><td colspan="8">Information Field
(Payload)</td></tr>
</table>

Figure 22.3-1. UNI cell format.

<table>
<tr><th></th><th>Bits→
Bytes↓</th><th>8</th><th>7</th><th>6</th><th>5</th><th>4</th><th>3</th><th>2</th><th>1</th></tr>
<tr><td rowspan="5">Header</td><td>1</td><td colspan="8">VPI</td></tr>
<tr><td>2</td><td colspan="4">VPI</td><td colspan="4">VCI</td></tr>
<tr><td>3</td><td colspan="8">VCI</td></tr>
<tr><td>4</td><td colspan="4">VCI</td><td colspan="3">PT</td><td>CLP</td></tr>
<tr><td>5</td><td colspan="8">HEC</td></tr>
<tr><td>Payload</td><td>6
7
...
53</td><td colspan="8">Information Field
(Payload)</td></tr>
</table>

Figure 22.3-2. NNI cell format.

Header Fields. The functions of the VPI and VCI fields have been discussed in Sections 22.1.1 and 22.1.2.

The Generic Flow Control (GFC) field, present only in the UNI cell format, is used to control cell flow from user to network. The GFC procedure is fairly complex, so we mention only that, from network to user, the first bit instructs the user to halt transmission (1) or not (0), and from user to network the last bit indicates if the terminal is controlled by GFC (1) or not (0). The other bits deal with the management of queues. We refer the reader to [14] for the details of the GFC procedure.

The Payload Type (PT) field indicates whether the payload contains OAM (first bit 1) or ATM-user-to-ATM-user (AAU) information (first bit 0). The other bits provide additional information, such as congestion.

The Cell Loss Priority (CLP) bit, when set to 1, indicates a lower priority cell that may be discarded in case of congestion; when set to 0 it indicates a high-priority cell for which every effort should be made to avoid discard. CLP can be set by the user or it can be set by traffic management processes to tag cells for possible discard when a user is exceeding contracted traffic values.

The Header Error Control (HEC) field is used for error checking on the contents of the header.

22.3.2 AAL/SAAL Layer Protocols

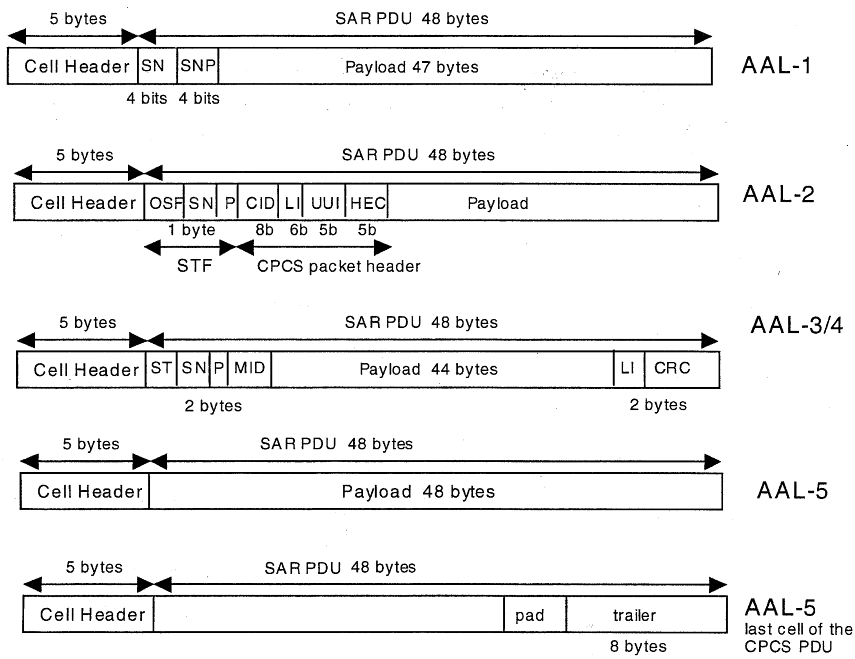
AAL Protocols. These protocols combine functions of the SAR and CPCS sublayers. They use SAR protocol data units (PDUs) to communicate with the ATM layer, which then transfers them across the ATM network in ATM cells. Different types of traffic require appropriately different treatment by the AAL layer, resulting in four types of AAL protocols:

- *AAL-1* provides *constant bit rate* (CBR) service and is used mainly for *circuit emulation services* (CESs) [15]. It requires timing synchronization between source and destination.
- *AAL-2* provides *variable bit rate* (VBR) real-time services for traffic that is carried by small packets, is bursty, and has low tolerance for delay (e.g., voice or video) [16]. It supports multiplexing of more than one stream in a single ATM connection.
- *AAL-3/4* is used for data services that are delay-tolerant [17]. It supports multiplexing, error detection, and sequence integrity. It is not widely used since the introduction of AAL5.
- *AAL-5* is a streamlined, low overhead version of AAL-3/4, typically used for IP over ATM, but fast becoming used for all types of services because of its efficiency [18]. This is the only protocol used in the control plane (i.e., for signaling).

During call setup the originating endpoint determines the type of AAL protocol to be used for the connection with information carried in the *AAL Type* field of the *AAL Parameters* information element (Fig. 22.5-2). The CS then takes the data payload from the source, encapsulates it into a CPCS PDU, and passes it to the SAR sublayer. The SAR sublayer divides it up and encapsulates it into SAR PDUs that fit into the 48-bit payload of the ATM-layer cell. The SAR PDU header uses some of the 48 bits from the cell payload. Figure 22.3-3 shows the SAR PDUs for the four AAL types, the overhead introduced, and the available payload.

With AAL-2, the data from the CPCS sublayer is packaged into CPCS PDUs that are typically small enough so that several of them can fit into one cell. Each CPCS PDU has a 3-byte header. Each SAR PDU has a 1-byte header (STF) that allows the delineation of the smaller CPCS packets. As a result, AAL-2 is very efficient for small units of data (less than 45 bytes).

With AAL-5, the data unit from the sending application is encapsulated by the CS sublayer into a CPCS PDU with an 8-byte trailer, and then padded to ensure it is a multiple of 48 bytes. The SAR sublayer then segments the CPCS PDU into 48-byte units that are transferred in cells with no overhead except for the last cell, which carries the padding and the CPCS trailer (to ensure sequence integrity). AAL-5 is the most efficient and widely used of the adaptation layer protocols, particularly for sending large data units.



STF: Start Field OSF: Offset Field SN: Sequence Number SNP: Sequence number protection
P: Parity (in AAL-2) P: Priority (in AAL3/4) CID: Connection Identifier LI: Length Indicator
UUI: User-to-User Indicator HEC: Header Error Control ST: Segment Type MID: Multiplexing Identification

Figure 22.3-3. SAR PDUs for the AAL protocols.

SSCOP. This protocol is used only for signaling (i.e., only in SAAL). Its main function is to provide reliable data transport between peer signaling entities, relieving the applications of the burden of checking for lost, corrupted, or misaligned messages. The protocol supports two modes of operation:

- Acknowledged
- Unacknowledged

The two modes of operation provide functions similar to the services offered by TCP and UDP (respectively) in IP networks.

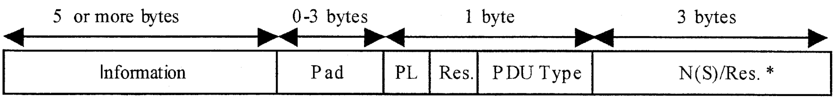
In acknowledged mode, SSCOP establishes a logical connection that ensures reception and sequence integrity of message by providing SDU sequence numbers and retransmission in case of error. Acknowledged mode is the standard mode used by signaling protocols.

Unacknowledged mode does not involve a logical connection, does not support retransmission, and does not guarantee reception of messages. It can be used optionally for management of semipermanent connections [19].

SSCOP uses primitives (called "signals" in ITU-T recommendations) to communicate with the signaling application and exchanges SSCOP PDUs with the ATM

TABLE 22.3-1 SSCOP PDUs

PDU	Acronym	Function
Begin	BGN	Establish connection
Begin Acknowledge	BGAK	Accept connection request
End	END	Release connection
End Acknowledge	ENDAK	Confirm connection release
Sequenced Data	SD	Sequentially numbered PDUs
Unnumbered Data	UD	Unsequenced PDUs



Res.: Reserved PL: Pad Length N(S): Send State Number (sequence number of the next PDU)
*: The last three bytes of the PDU contain N(S) in SD PDUs and is not used (reserved) in UD PDUs

Figure 22.3-4. SD and UD PDU formats.

layer. The main PDU types are listed in Table 22.3-1. BGN, BGAK, END, and ENAK are link management functions, used to establish and release a SSCOP (signaling) connection between two nodes. Once established, a SSCOP connection is semipermanent and is used to transfer messages for all calls managed through a signaling virtual channel (SVC—see Section 22.4).

SD is the PDU for acknowledged data transfer; UD is the PDU for unacknowledged data transfer. A signaling message is carried as the information field of a SD PDU. The formats of SD and UD PDUs are shown in Fig. 22.3-4. For other PDUs, we refer the reader to [11].

22.4 INTRODUCTION TO ATM SIGNALING

Signaling in an ATM network is used to establish, manage, and release VCCs and VPCs, as well as to negotiate QoS and other data transfer parameters that are used during the connection. Several ATM signaling protocols exist, and they can be grouped into two classes based on the type of interface: UNI or NNI. They can also be categorized in terms of sponsorship: ITU-T or ATM Forum. Table 22.4-1 provides a summary of the most commonly used ATM signaling protocols.

TABLE 22.4-1 ATM Signaling Protocols

Interface Type	ITU-T Protocol	ATM Forum Protocol
UNI	DSS2	UNI (V4)
NNI	B-ISUP	PNNI AINI B-ICI

Signaling messages are transported in dedicated, semipermanent virtual channel connections called *signaling virtual channels* (SVCs), which must be preestablished for signaling communication to take place. A signaling virtual channel is identified by the default values VPI = 0 and VCI = 5, but different VPI and VCI values can be set by provisioning or through the meta-signaling protocol, not described in this book [20]. Meta-signaling, when used, sends its messages on a permanent VCC, called the meta-signaling channel, with default values of VPI = 0 and VCI = 0.

Two arrangements are supported for SVCs:

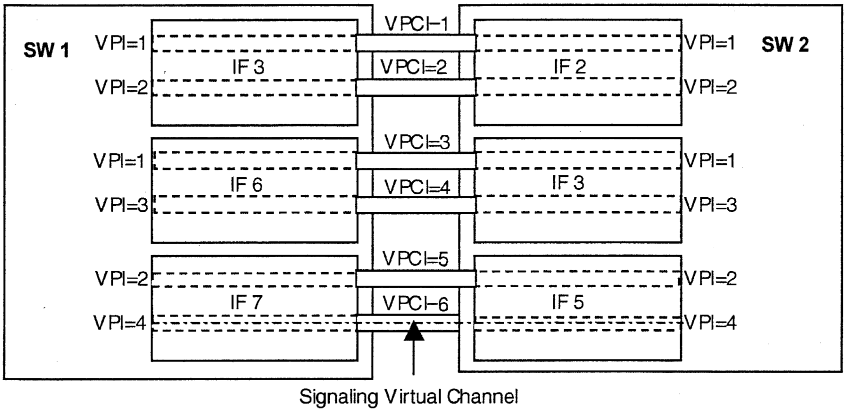
1. Associated signaling
2. Nonassociated signaling

With associated signaling, a signaling virtual channel controls only the virtual channel connections in its own VPC. With nonassociated signaling, a SVC controls virtual channel connections in its own VPC as well as in other VPCs. By default, a signaling virtual channel operates in nonassociated mode and controls all the VCCs and VPCs in its physical link, except the ones that are controlled by associated signaling, if any. It is also possible to set up a SVC to control VPCs and VCCs in multiple physical links between the same two nodes. To provide maximum flexibility, signaling messages use *virtual path connection identifiers* (VPCI) instead of VPIs to identify VPCs. The VPCI decouples a signaling virtual channel from its physical link, allowing it to address VPCs regardless of the links they are in. Its 16-bit field can address a larger number of VPCs than the VPI. The use of VPCI makes possible the insertion of cross-connects in front of ATM switches, a configuration that can cause VPCs from different physical links to be connected to the same switch interface.

The two parameters that signaling messages use to identify a virtual channel connection are thus VCI and VPCI: VCI is the same identifier used in the cell header, while VPCI is used only by signaling and is mapped by each switch into the VPI used in the cell header. The mapping is based on tables that are configured by provisioning. VPCI values have only local significance and are unique only within the domain of a particular SVC. A conceptual example of VPCI-to-VPI mapping is shown in Fig. 22.4-1. In the example, messages in the SVC (from SW 1) identify VPI = 1 on IF 3 as VPCI = 1, VPI = 1 on IF 6 as VPCI = 3, and so on.

With the UNI and PNNI protocols, the VPCI/VCI assignment is controlled by the succeeding (downstream) side, but specific values can be requested by the preceding side. Values are carried in the Connection Identifier IE (see Table 22.6-1).

With the B-ISUP and B-ICI protocols, each switch controls half of the VPCI values (odd or even numbers) used by a SVC. The switch with the highest signaling point code controls the even numbers. The originating switch always starts as the assigning switch and defaults to nonassigning only if it does not have the necessary resources in its VPCI pool. The assigning switch sends the chosen VPCI/VCI values in the CEI information element of the IAM or IAA message (see Table 22.7-2). When nonassigning, a switch omits the CEI from the message.



SW n: ATM Switch IF: physical interface

Figure 22.4-1. Conceptual example of VPCI to VPI mapping.

Lower-Layer Protocols. Signaling, as a higher-layer application, relies on the services of lower-layer ATM protocols (Fig. 22.2-1) to perform its functions, namely:

- ATM layer (cells)
- AAL-5 protocol
- SCCOP

all described in Section 22.3.

22.5 SIGNALING AT THE UNI INTERFACE

Two signaling protocols are defined for the UNI interface:

1. DSS2 (digital signaling system No. 2) specified in ITU-T Rec. Q.2931 [19]
2. UNI signaling specified by the ATM Forum [21,22]

DSS2 is intended for public UNI interfaces and is built and based on narrowband ISDN interfaces DSS-1 signaling (Rec. Q.931—Chapter 10). AF UNI signaling is intended for both private and public interfaces and its specification is written as a delta on Q.2931. The differences between the two are relatively minor and appear only at a fairly low level of detail.

Connection Control Messages. These messages are shown in Table 22.5-1. The main difference with the DSS1 protocol is the absence of the DISCONNECT message.

General Message Format. A UNI message is composed of a series of information elements (IEs), some of fixed length and some of variable length.

TABLE 22.5-1 UNI Connection Control Messages

Message Type	Cg ↔ NW	NW ↔ Cd	DSS1[a]	Function
ALERTING	←	←		Called party alerted
CALL PROCEEDING	←	←		Connection request is being processed
CONNECT	←	←		Called party accepts connection
CONNECT ACKNOWLEDGE	→	→		Connection established
NOTIFY	↔	↔		Report information
RELEASE	↔	↔		Request to clear connection
RELEASE COMPLETE	→	←		Connection released
RESTART	↔	↔		Restart VCC or VPC
RESTART ACKNOWLEDGE	↔	↔		Restart of VCC or VCP complete
SETUP	→	→		Start connection
STATUS	↔	↔		Report error conditions
STATUS ENQUIRY	↔	↔		Request STATUS message

[a]Common with the DSS1 protocol.

Note: NW, network; Cg, calling endpoint; Cd, called endpoint.

Fixed-length IEs are mandatory for all message types. Variable-length IEs can be optional or they can be mandatory for some message types.

The format for the mandatory IEs is shown in Fig. 22.5-1. The *protocol discriminator* identifies the signaling protocol. The *call reference value* identifies the connection at the UNI interface and is unique within individual signaling virtual channels. The *message type* identifies the function of the message: the first byte identifies the message type and the second contains spare bits and flags. The *message length* indicates the number of bytes in the variable part of the message.

Variable-length IEs depend on the type of message and the most important ones are listed in Table 22.5-2. *AAL parameters* conveys the AAL type (AAL-1

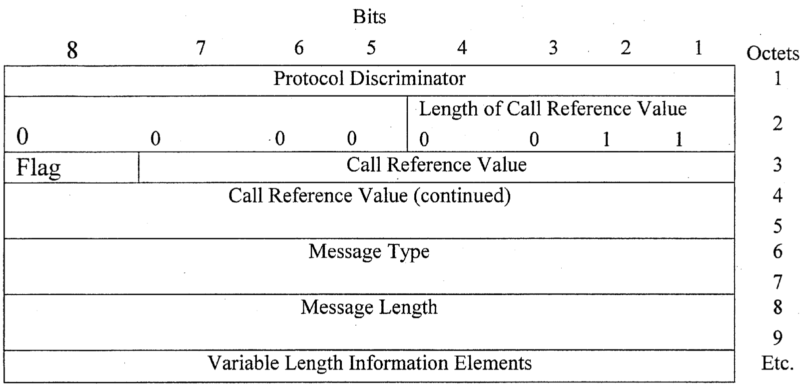


Figure 22.5-1. General message organization example. (From Rec. Q.2931. Reproduced with the kind permission of ITU.)

TABLE 22.5-2 Variable-Length Information Elements in UNI Signaling Messages

Message Type → Information Element ↓	DSS1[a]	ALERTING	CALL PROCEEDING	CONNECT	CONNECT ACKNOWLEDGE	NOTIFY	RELEASE	RELEASE COMPLETE	RESTART	RESTART ACKNOWLEDGE	SETUP	STATUS	STATUS ENQ[c]
AAL parameters				O							O		
ATM traffic descriptor											M		
Called party number											O		
Calling party number											O		
Cause							M	O				M	
Connection identifier		O	O	O					O	O	O		
Notification Indicator	[b]	O	O	O	O	M	O				O		
QoS parameter											M		
Restart indicator	[b]								M	M			

[a]Common with the DSS1 protocol.
[b]Not described in Chapter 10.
[c]No variable-length IE in this message.

Bits→ 8	7	6	5	4	3	2	1	Octets↓
AAL Parameters IE identifier								
0	1	0	1	1	0	0	0	1
1	Coding standard		IE Instruction Field					2
			Flag	Reserved	IE action ind.			
Length of AAL parameter content								3 4
AAL Type								5
Further content								6 ....

Figure 22.5-2. AAL Parameters information element. (From Rec. Q.2931. Reproduced with the kind permission of ITU.)

through 5). *ATM traffic descriptor* contains the set of traffic parameters requested for the call, such as peak cell rate. *Notification indicator* conveys call-related information. *QoS parameter* contains the QoS class [5]. *Restart indicator* indicates the virtual channel, or the range thereof, that needs to be restarted. The other IEs are common with the DSS1 protocol, and for them we refer the reader to Chapter 10.

An example of a variable-length IE format is shown in Fig. 22.5-2. A detailed explanation of all the fields is found in [19].

22.5.1 UNI Signaling Sequence

An example of a call sequence is shown in Fig. 22.5-3. The originating ATM endpoint starts a connection with a Setup message sent to the ATM switch it is connected to. The Setup message carries the requested QoS parameters. When the ATM network receives the Setup message, it can optionally send a Call Proceeding message back to the endpoint, while the connection request is processed. This is useful in large networks where it may take time for the response from the destination to reach the originating endpoint, and timeouts may abort the call at the source. The network sends the Setup message to the destination endpoint, which also may optionally respond with a Call Proceeding message while the request is processed. The destination endpoint may send an Alerting message to the network to inform it that the called party is being alerted; the network then forwards the Alerting message to the originating endpoint. When the destination endpoint accepts the connection request, it sends a Connect message to the network, which responds with a Connect Acknowledge while forwarding the Connect to the originating endpoint. Upon receipt of the Connect, the originating endpoint responds with a Connect Acknowledge and the ATM data flow between the users starts. Either the source or the destination can terminate the connection by sending a Release message, responded to by the network with the release of the resources used for the connection and with a Release Complete. The network forwards the Release message to the originating or the destination endpoint, which responds with a Release Complete.

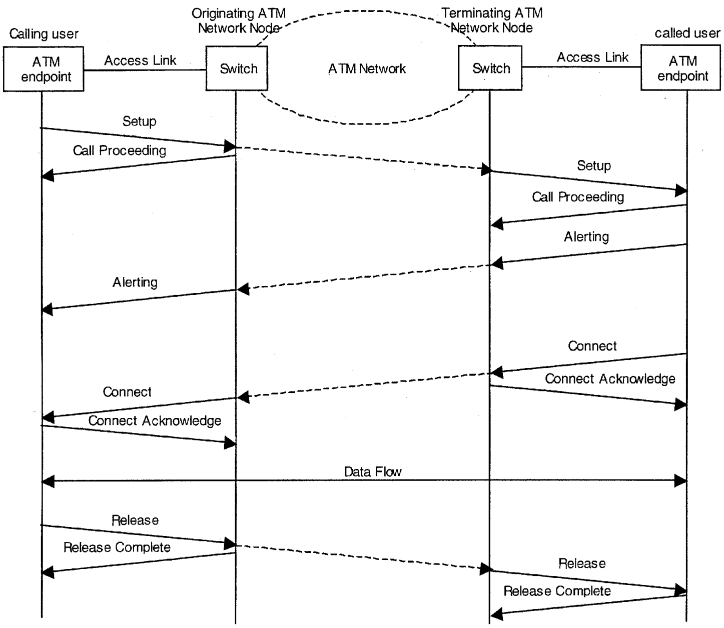


Figure 22.5-3. Example of UNI signaling sequence.

22.6 THE PNNI PROTOCOL

The PNNI (Private Network–Network Interface) protocol is an ATM Forum standard designed for network interfaces between private ATM networks or between ATM nodes within private networks [23]. The specification also allows use of the protocol between public ATM networks.

PNNI specifications call the upstream node *preceding* and the downstream node *succeeding*. Signaling messages that are forwarded to the next node after being received and processed are said to have *global significance*, while messages with *local significance* are not forwarded.

The PNNI protocol is composed of two parts:

1. Routing protocol
2. Signaling protocol

22.6.1 PNNI Routing Protocol

Before addressing the signaling part of the PNNI protocol, a few words about the PNNI routing protocol are needed, since it plays a role that affects connection setup. Routing protocols are a basic function of packet networks and are used by Layer 3 nodes, such as routers, to exchange reachability and link status information without manual intervention. That exchange of information is done so that a route (path) from the source to the destination can be determined at all times, as nodes and links go in and out of service. Based on the information received, each L3 node/router generates a *routing table*, which is updated dynamically. This is in contrast to the essentially manual management of routing information in a circuit-switched network. There are different families of routing protocols, and the PNNI routing protocol belongs to the family called *link-state*. With link-state protocols, nodes exchange information about all their links and use that information to build a *topological database* containing an image of the network. A *routing table* is then derived from the topological database by applying an algorithm called the *Dijkstra algorithm* (from the name of its creator [23]).

The PNNI routing protocol works in concert with ATM's QoS and traffic management capabilities to choose the appropriate path to the destination. A procedure, called *generic call admission control* (GCAC), is used for choosing one of the available paths from the routing table, factoring-in several parameters: Destination ATM address, Traffic Class, Traffic Contract, QoS, Link Constraints.

PNNI uses *source routing*, where the originating ATM node determines the total path to the destination. IP networks, on the other hand, use *hop-by-hop routing*, in which each individual node determines the next hop toward the destination. An ATM switch uses GCAC to generate a list, called the *designated transit list* (DTL), specifying the nodes that need to be traversed to reach the destination. The DTL is sent forward in the DTL information element of the Setup message (Table 22.6-1). Because of the hierarchical nature of PNNI routing, the DTL, as generated by the source, can be a summary route that is expanded as needed by intermediate nodes. For the details of the PNNI routing protocol we refer the reader to [23].

22.6.2 PNNI Signaling Protocol

The PNNI signaling protocol is based on UNI V4.1, which makes it based on DSS2. In addition to the functionality in DSS2, PNNI includes a crankback function (see Section 1.3.4), which it shares with AINI and B-ISUP (discussed in later sections).

PNNI Connection Control Messages. The main connection control messages are the same as the UNI messages in Table 22.5-1, with the exception of the Connect Acknowledge message, which is not used. The general message format is the same as for DSS2 (Fig. 22.5-1), with a different binary code (1 1 1 1 0 0 0 0) for the Protocol Discriminator. The most important variable-length IEs are listed in Table 22.6-1. *Crankback* indicates that the crankback procedure (see next section) has been invoked and conveys the cause for it. *Designated transit list* contains the DTL (Section 22.6.1).

TABLE 22.6-1 Variable-Length Information Elements in PNNI Signaling Messages

Message Type → Information Element ↓	UNI[a]	ALERTING	CALL PROCEED	CONNECT	NOTIFY	RELEASE	RELEASE COMPL	RESTART	RESTART ACK	SETUP	STATUS	STATUS ENQ[b]
AAL parameters				O						O		
ATM traffic descriptor				O						M		
Called party number										M		
Calling party number										O		
Cause						M	O				M	
Connection identifier			M					O	O	O		
Crankback						O	O					
Designated transit list (DTL)										M		
Notification indicator		O		O	M	O				O		
QoS parameter										O		
Restart indicator								M	M			

[a]Common with UNI signaling.
[b]No variable-length IEs in this message.

22.6.3 PNNI Signaling Sequence

A connection is started with a Setup message sent by a preceding node, along the path specified by the DTL as explained in Section 22.6.1. The Setup message includes the DTL as well as the requested QoS parameters. When the node on the succeeding side receives the Setup message, it applies the GCAC algorithm again, to determine if it can meet the QoS requirements. If so, the appropriate resources are reserved, VCI and/or VPI values are assigned, and the call is forwarded to the next node. If not (or if the call cannot be routed), the crankback procedure is triggered (see below). When all information needed for call setup has been received, the succeeding node sends a Call Proceeding message to the preceding node to inform it that the call is being processed. When the destination node determines that the called party is being alerted, it sends an Alerting message in the backward direction. The Alerting is forwarded from succeeding to preceding node up to the originating node. Similarly, when the called party accepts the call, a Connect message is sent backward from the destination node and forwarded from succeeding to preceding node up to the originating node, after which the path is established and the flow of user data can begin. Either the origination or the destination node can terminate the connection. Termination of the connection starts with a Release message sent by either the preceding node (for source-originated termination) or the succeeding node (for destination-originated termination), responded to with Release Complete when the resources used for the call are released. An example of a call sequence is shown in Fig. 22.6-1.

Crankback. This procedure is started by sending a clearing message, such as Release or Release Complete with the crankback IE in it, to the preceding node.

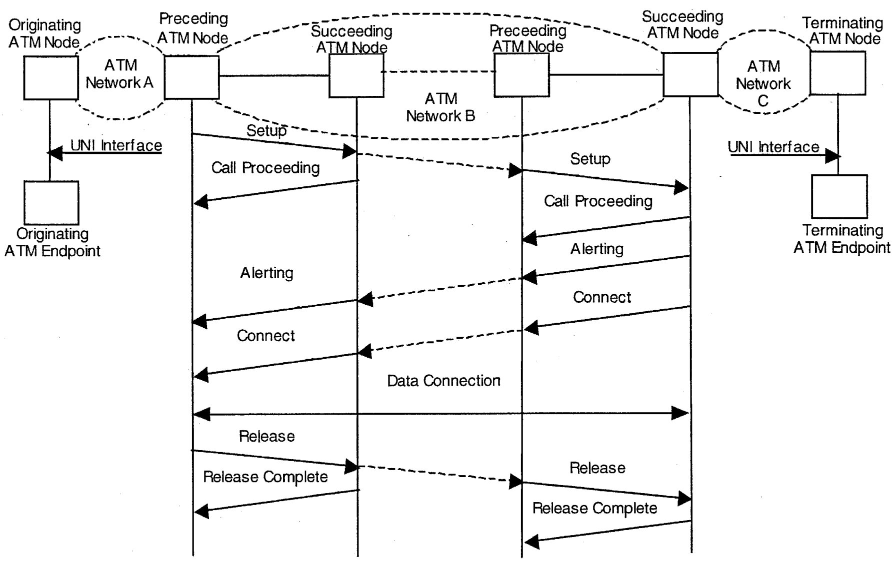


Figure 22.6-1. Example of a PNNI call sequence.

When the preceding node receives the message, it tries an alternate route. The procedure is useful because network status can change dynamically and delays in the propagation of routing protocol messages may affect the accuracy of the routing table used by a node.

PNNI Point-to-Multipoint Connections. In addition to the standard point--to-point connections, PNNI (like UNI) supports point-to-multipoint connections. PNNI follows, with few modifications, ITU-T Rec. Q.2971 [6].

22.7 THE B-ISUP SIGNALING PROTOCOL

Broadband ISDN User Part, or B-ISUP, is the ITU-T signaling protocol for public NNI [24–27]. Its main target is international connections, but use in national networks is allowed. As the name implies, B-ISUP is based on and derived from the ISUP protocol, described in Chapter 11. Like ISUP, B-ISUP uses the services of MTP-3, but as amended by Rec. Q.2210 [28], which defines what is known informally as MTP-3b. MTP-3b uses the services of SAAL (SSCF) at its lower boundary, instead of MTP-2, and interfaces with B-ISUP at its upper boundary. MTP-3b uses the same primitives as MTP-3 to exchange parameters with B-ISUP. The protocol stack is shown in Fig. 22.7-1.

The B-ISUP message format, shown in Fig. 22.7-2, is structured with a header followed by the *message content*. The *routing label* (RL) is, strictly speaking, part of MTP-3b and is defined the same way as for MTP-3 (Chapters 7 and 8). RL contains the destination point code, the originating point code, and the signaling link selection fields. The *message type code* field is the same as for ISUP and most of its values are also the same as for ISUP.

Message compatibility information notifies the receiving side of what to do if a message is not understood. *Message content* is a sequence of parameters, whose general format is shown in Fig. 22.7-3. A difference with ISUP is that the CIC field is not present, since ATM uses VCI and VPI to identify a user-to-user connection. VCI and VPCI (the proxy for VPI in signaling messages) are carried in parameters such as the *connection element identifier* (Fig. 22.7-5).

The most significant B-ISUP message types are shown in Table 22.7-1. The main parameters for call-control messages are shown in Table 22.7-2. B-ISUP

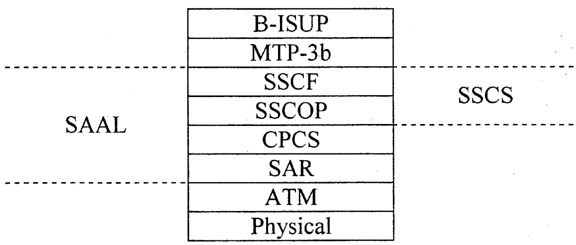


Figure 22.7-1. B-ISUP protocol stack.

Bits 8 7 6 5 4 3 2 1	
Routing Label	5 Octets
Message Type Code	1 Octet
Message Length	2 Octets
Message Compatibility Information	1 Octet
Message Content	Variable length

Figure 22.7-2. B-ISUP message format. (From Rec. Q.2763. Reproduced with the kind permission of ITU.)

Bits 8 7 6 5 4 3 2 1	
Parameter Name	1 octet
Length Indicator	2 octets
Parameter Compatibility Information	1 octet
Parameter Content	Variable octet length

Figure 22.7-3. B-ISUP parameter format. (From Rec. Q.2763. Reproduced with the kind permission of ITU.)

TABLE 22.7-1 B-ISUP Call-Control Messages

Acronym	Name	ISUP[a]	Function
ACM	Address Complete		Complete called party number received
ANM	Answer		Called party has answered
CPG	Call Progress		Call-related event at the called side
IAA	IAM Acknowledge		Request for resources is accepted
IAM	Initial Address		Start connection
IAR	IAM Reject		Request for resources is rejected
REL	Release		Release connection
RES	Resume		Resume communication after a suspend
RLC	Release Complete		Resources have been released
SUS	Suspend		Interrupt communication w/o releasing the connection
USR	User-to-User Information		Exchange information between users

[a]Common with ISUP.

TABLE 22.7-2 Parameters in B-ISUP Call-Control Messages

Message Type → Parameter name ↓	ISUP[a]	ACM	ANM	COA	CPG	IAA	IAM	IAR	REL	RES	RLC	SUS	USR
AAL parameters			x				x						
ATM cell rate			x				x						
Automatic congestion level								x	x				
Automatic rerouting (crankback)							x		x				
Called party number							x						
Calling party number							x						
Cause indicators		x			x			x	x		x		
Connection element identifier (CEI)						x	x						
Destination signalling identifier		x	x	x	x	x		x	x	x	x	x	x
Priority							x						
Quality of service (QoS)							x						
Redirection number		x	x		x				x				
Suspend/resume indicators										x		x	
User-to-user information		x	x		x		x		x				x

[a]Common with ISUP.

specifications do not label parameters as mandatory or optional and those characteristics can be deduced only from the details of call-control procedures. Two examples of format for the "Parameter Content" field of Figs. 22.7-3 are shown in Figs. 22.7-4 and 22.7-5. For a complete list and the detailed definitions, we refer the reader to [26].

An example of a B-ISUP call-setup sequence, where the UNI portions of the connection use DSS2 signaling, is shown in Fig 22.7-6 [29].

Bits

8	7	6	5	4	3	2	1	Octets
Odd/Even	Nature of Address Indicator							1
Internal Network No. Ind.	Numbering Plan Indicator			Spare				2
2nd address signal				1st address signal				3
								……..
Filler (if necessary)				Nth address signal				n

Figure 22.7-4. Called Party Number parameter field. (From Rec. Q.2763. Reproduced with the kind permission of ITU.)

Bits 8 7 6 5 4 3 2 1	Octets
Virtual Path Connection Identifier (VPCI)	1 2
Virtual Channel Identifier (VCI)	3 4

Figure 22.7-5. Connection Element Identifier parameter field. (From Rec. Q.2763. Reproduced with the kind permission of ITU.)

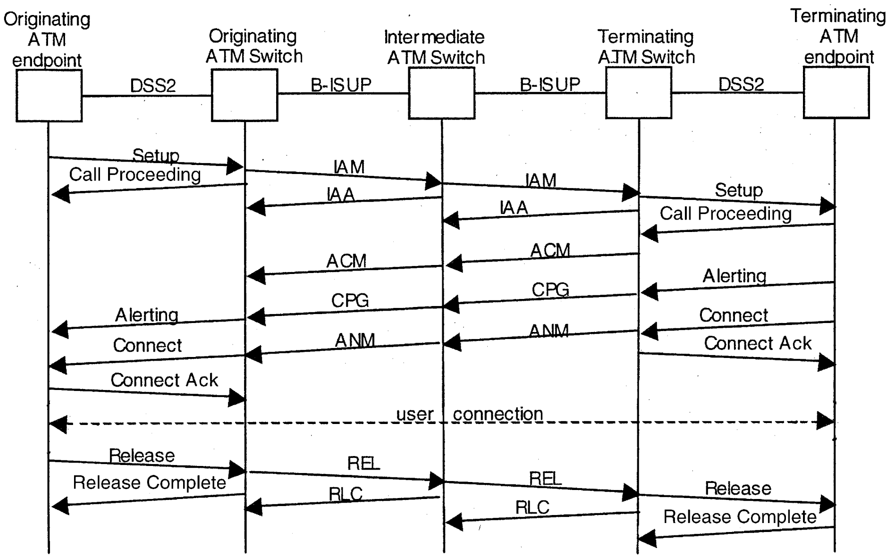


Figure 22.7-6. Example of B-ISUP call setup sequence with DSS-2 interworking.

22.8 OTHER NNI SIGNALING PROTOCOLS

AINI (ATM Inter-Network Interface). AINI is an ATM Forum signaling protocol specifically designed for connections between ATM networks that use PNNI internally [30]. AINI is based on the PNNI protocol with very few modifications: the designated transit list (DTL) is not supported, and a few information elements and a few messages are not supported (all the messages listed in Table 22.5-1 are supported). PNNI signaling messages and IEs that are of local significance are not transported across the AINI; all other PNNI messages and IEs are mapped into corresponding AINI messages and IEs.

B-ICI (B-ISDN Inter-Carrier Interface). The B-ICI [31,32] is AF's NNI signaling protocol between public networks managed by different telecom administrations. The signaling portion of the B-ICI specification is based on the B-ISUP recommendations, with only minor differences. In addition to signaling, and unlike B-ISUP, B-ICI includes detailed requirements for physical, data link, and network layers.

22.9 ATM ADDRESSING

ATM networks use two types of addresses for connection setup:

1. Telephony numbering plan addresses, specified by ITU-T in Rec. E.164 (Section 1.2). This type of address is called *native E.164*.
2. *ATM end system addresses* (AESAs), described in the *Addressing Reference Guide* and in the *User Guide* of the ATM Forum [33,34].

AESA are based on ISO's *network service access point* (NSAP) addresses, with some very minor differences [35]. For that reason they are sometimes referred to as *NSAP AESAs* or just *NSAP addresses*. The ATM Forum specifications define four types of AESAs, all having a fixed length of 20 octets:

1. DCC (data country code) format
2. ICD (international code designator) format
3. E.164 format (AESA)
4. Local AESA format

DCC addressing authority rests with the ISO, who assigns a national code to each country, per ISO 3166, and delegates authority for the *domain-specific part*

20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
AFI	DCC		HO-DSP										ESI						SEL
←IDP→			←DSP→																
	←IDI→																		

DCC AESA Format

20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
AFI	E.164								HO-DSP				ESI						SEL
←IDP→									←DSP→										
	←IDI→																		

E.164 AESA Format

AFI: Authority and Format Indicator
ESI: End-System Identifier
IDI: Initial Domain Identifier
DCC: Data Country Code
HO-DSP: High-Order Domain Specific Part
IDP: Initial Domain Part
DSP: Domain Specific Part
SEL: Selector

Figure 22.9-1. AESA addressing formats. (Courtesy of the ATM Forum.)

(DSP) of the address to national entities (e.g., ANSI for USA, FEI for UK, DIN for FRG, etc.).

ICD addressing authority rests with EDIRA (*Electronic Data Interchange Registration Authority*), an international nonprofit organization connected with the ISO. EDIRA assigns codes to entities that are then free to define specific coding schemes for use by industry (such as the bar code system). The entity owning the code point defines the domain-specific part of the address.

The E.164 format (AESA) is the standard telephony numbering plan address embedded in an AESA 20-octet format. It is also called *embedded E.164* to distinguish it from the native form mentioned earlier.

The Local AESA format is rarely used.

Figure 22.9-1 shows the DCC and the E.164 AESA (embedded) formats.

22.10 ACRONYMS

AAL	ATM adaptation layer
AAU	ATM user to ATM user
AESA	ATM end system address
AF	ATM Forum
AINI	ATM Inter-Network Interface
ATM	Asynchronous transfer mode
B-ICI	B-ISDN Inter-Carrier Interface
B-ISDN	Broadband ISDN
B-ISUP	Broadband ISDN User Part
BSS	Broadband switching system
CBR	Constant bit rate
CES	Circuit emulation service
CLP	Cell loss priority
CPCS	Common part CS
CS	Convergence sublayer
DCC	Data country code
DSP	Domain Specific Part
DSS2	Digital Signaling System No. 2
DTL	Designated transit list
EDIRA	Electronic Data Interchange Registration Authority
GCAC	Generic call admission control
GFC	Generic flow control
HEC	Header error control
ICD	International code designator

IE	Information element
IETF	Internet Engineering Task Force
ISUP	ISDN User Part
ITU	International Telecommunication Union
ITU-T	Telecommunication Standardization Sector of ITU
LAN	Local area network
MFA	MPLS, Frame Relay, ATM
MPLS	Multiprotocol label switching
MTP	Message Transfer Part
NNI	Network–network interface or network–node interface
PDU	Protocol data unit
PNNI	Private Network–Network Interface Protocol
PT	Payload type
PVCC	Permanent virtual channel connection
PVPC	Permanent virtual path connection
QoS	Quality of service
SAAL	Signaling ATM adaptation layer
SAR	Segmentation and reassembly
SDH	Synchronous Data Hierarchy
SONET	Synchronous Optical Network
SSCF	Service-Specific Coordination Function
SSCOP	Service-Specific Connection-Oriented Protocol
SSCS	Service-specific CS
STF	Start field
SVC	Signaling virtual channel
SVCC	Switched virtual channel connection
SVPC	Switched virtual path connection
TC	Transmission Convergence
TDM	Time-division multiplexing
UNI	User–network interface
VBR	Variable bit rate
VCC	Virtual channel connection
VCI	Virtual channel identifier
VCL	Virtual channel link
VPC	Virtual path connection
VPCI	Virtual path connection identifier
VPI	Virtual path identifier
VPL	Virtual path link
WAN	Wide area network

22.11 REFERENCES

1. William Stallings, *Data and Computer Communications*, Prentice Hall, Upper Saddle River, NJ, 1997.
2. Timothy Kwok, *ATM: The New Paradigm for Internet, Intranet & Residential Broadband Services and Applications*, Prentice Hall, Upper Saddle River, NJ, 1998.
3. *B-ISDN ATM Layer Cell Transfer Performance*, Rec. I.356, ITU-T, Geneva, Mar. 2000.
4. *Traffic Control and Congestion Control in B-ISDN*, Rec. I.371, ITU-T, Geneva, Mar. 2000.
5. *Traffic Management Specification Version 4.1*, AF-TM-0121.000, ATM Forum, Mar. 1999.
6. *Broadband Integrated Services Digital Network (B-ISDN)—Digital Subscriber Signaling System No. 2 (DSS2) User–Network Interface Layer 3 Specification for Point-to-Multipoint Call/Connection Control*, Rec. Q.2971, ITU-T, Geneva, Oct. 1995.
7. *B-ISDN Protocol Reference Model and its Application*, Rec. I.321, ITU-T, Geneva, Apr. 1991.
8. *B-ISDN Signaling ATM Adaptation Layer (SAAL—Overview Description*, Rec. Q.2100, ITU-T, Geneva, July 1994.
9. *B-ISDN Signalling ATM Adaptation Layer—Layer Management for the SAAL at the Network Node Interface (NNI)*, Rec. Q.2144, ITU-T, Geneva, Oct. 1995.
10. *Internetworking Technologies Handbook*, 3rd ed., Cisco Press, 2001.
11. *B-ISDN ATM Adaptation Layer—Service Specific Connection Oriented Protocol (SSCOP)*, Rec. Q.2110, ITU-T, Geneva, July 1994.
12. *B-ISDN Signalling ATM Adaptation Layer—Service Specific Coordination Function for Support of Signalling at the User–Network Interface (SSCF at UNI,)* Rec. Q.2130, ITU-T, Geneva, July 1994.
13. *B-ISDN ATM Adaptation Layer—Service Specific Coordination Function for Signalling at the Network Node Interface (SSCF at NNI)*, Rec. Q.2140, ITU-T, Geneva, Feb. 1995.
14. *B-ISDN ATM Layer Specification*, Rec. I.361, ITU-T, Geneva, Feb. 1999.
15. *B-ISDN ATM Adaptation Layer Specification: Type 1 AAL*, Rec. I.363.1, ITU-T, Geneva, Aug. 1996.
16. *B-ISDN ATM Adaptation Layer Specification: Type 2 AAL*, Rec. I.363.2, ITU-T, Geneva, Nov. 2000.
17. *B-ISDN ATM Adaptation Layer Specification: Type 3/4 AAL*, Rec. I.363.3, ITU-T, Geneva, Aug. 1996.
18. *B-ISDN ATM Adaptation Layer Specification: Type 5 AAL*, Rec. I.363.5, ITU-T, Geneva, Aug. 1996.
19. *Digital Subscriber Signaling System No. 2—User–Network Interface (UNI) Layer 3 Specification for Basic Call/Connection Control*, Rec. Q.2931, ITU-T, Geneva, Feb. 1995.
20. *B-ISDN Meta-signalling Protocol*, Rec. Q.2120, ITU-T, Geneva, Feb. 1995.
21. *ATM User–Network Interface (UNI) Signaling Specification Version 4.1*, AF-SIG-0061.002, ATM Forum, Apr. 2002.
22. *ATM User–Network Interface (UNI) Specification Version 4.1*, AF-ARCH-0193.000, ATM Forum, Apr. 2002.

23. *Private Network–Network Interface Specification Version 1.1 (PNNI 1.1)*, AF-PNNI-0055.002, ATM Forum, Apr. 2002.
24. *Functional Description of the B-ISDN User Part (B-ISUP) of Signalling System No. 7*, Rec. Q.2761, ITU-T, Geneva, Dec. 1999.
25. *General Functions of Messages and Signals of the B-ISDN User Part (B-ISUP) of Signalling System No. 7*, Rec. Q.2762, ITU-T, Geneva, Dec. 1999.
26. *Signalling System No. 7 B-ISDN User Part (B-ISUP)—Formats and Codes*, Rec. Q.2763, ITU-T, Geneva, Dec. 1999.
27. *Signalling System No. 7 B-ISDN User Part (B-ISUP)—Basic Call Procedures*, Rec. Q.2764, ITU-T, Geneva, Dec. 1999.
28. *Message Transfer Part Level 3—Functions and Messages Using the Services of ITU-T Recommendation Q.2140*, Rec. Q.2210, ITU-T, Geneva, July 1996.
29. *Interworking Between Signalling System No. 7 Broadband ISDN User Part (B-ISUP) and Digital Subscriber Signalling System No. 2 (DSS2)*, Rec. Q.2650, ITU-T, Geneva, Dec. 1999.
30. *ATM Inter-Network Interface Specification Version 1.1 (AINI 1.1)*, AF-CS-0125.002, ATM Forum, Sept. 2002.
31. *BISDN Inter Carrier Interface (B-ICI) Specification Version 2.0 (Integrated)*, AF-BICI-0013.003, ATM Forum, Dec. 1995.
32. *Addendum to BISDN Inter Carrier Interface (B-ICI) Specification, v2.0 (B-ICI Specification, v2.1)*, AF-BICI-0068.000, ATM Forum, Nov. 1996.
33. *ATM Forum Addressing User Guide Version 1.0*, AF-RA-0105.000, ATM Forum, Jan. 1999.
34. *ATM Forum Addressing Reference Guide*, AF-RA-0106.000, ATM Forum, Feb. 1999.
35. *Information Technology—Open Systems Interconnection—Network Service Definition*, Rec. X.213, ITU-T, Geneva, Oct. 2001.

INDEX

Signaling in Telecommunication Networks, Second Edition, by John G. van Bosse and Fabrizio Devetak

WILEY SERIES IN TELECOMMUNICATIONS AND SIGNAL PROCESSING

John G. Proakis, Editor
Northeastern University

Introduction to Digital Mobile Communications
Yoshihiko Akaiwa

Digital Telephony, 3rd Edition
John Bellamy

ADSL, VDSL, and Multicarrier Modulation
John A. C. Bingham

Biomedical Signal Processing and Signal Modeling
Eugene N. Bruce

Elements of Information Theory, 2nd Edition
Thomas M. Cover and Joy A. Thomas

Erbium-Doped Fiber Amplifiers: Device and System Developments
Emmanuel Desurvire

Fiber-Optic Systems for Telecommunications
Roger L Freeman

Fundamentals of Telecommunications, 2nd Edition
Roger L. Freeman

Practical Data Communications, 2nd Edition
Roger L. Freeman

Radio System Design for Telecommunications, 2nd Edition
Roger L. Freeman

Telecommunication System Engineering, 4th Edition
Roger L. Freeman

Telecommunications Transmission Handbook, 4th Edition
Roger L. Freeman

Introduction to Communications Engineering, 2nd Edition
Robert M. Gagliardi

Optical Communications, 2nd Edition
Robert M. Gagliardi and Sherman Karp

Efficient Algorithms for MPEG Video Compression
Dzung Tien Hoang and Jeffrey Scott Vitter

Active Noise Control Systems: Algorithms and DSP Implementations
Sen M. Kuo and Dennis R. Morgan

Mobile Communications Design Fundamentals, 2nd Edition
William C. Y. Lee

Robust Adaptive Beamforming
Jian Li and Petre Stoica, Editors

Expert System Applications for Telecommunications
Jay Liebowitz

Polynomial Signal Processing
V. John Mathews and Giovanni L. Sicuranza

Digital Signal Estimation
Robert J. Mammone, Editor

Digital Communication Receivers: Synchronization, Channel Estimation, and Signal Processing
Heinrich Meyr, Marc Moeneclaey, and Stefan A. Fechtel

Synchronization in Digital Communications, Volume I
Heinrich Meyr and Gerd Ascheid

Business Earth Stations for Telecommunications
Walter L. Morgan and Denis Rouffet

Wireless Information Networks, 2nd Edition
Kaveh Pahlavan and Allen H. Levesque

Satellite Communications: The First Quarter Century of Service
David W. E. Rees

Fundamentals of Telecommunication Networks
Tarek N. Saadawi, Mostafa Ammar, with Ahmed El Hakeem

Analogue and Digital Microwave Links: Theory and Design
Carlos Salema

Microwave Radio Links: From Theory to Design
Carlos Salema

Meteor Burst Communications: Theory and Practice
Donald L. Schilling, Editor

Digital Communication over Fading Channels, 2nd Edition
Marvin K. Simon and Mohamed-Slim Alouini

Digital Signal Processing: A Computer Science Perspective
Jonathan (Y) Stein

Vector Space Projections: A Numerical Approach to Signal and Image Processing, Neural Nets, and Optics
Henry stark and Yongyi Yang

Signaling in Telecommunication Networks, 2nd Edition
John G. Van Bosse and Fabrizio Devetak

Telecommunication Circuit Design, 2nd Edition
Patrick D. van der Puije

Worldwide Telecommunications Guide for the Business Manager
Walter H. Vignault