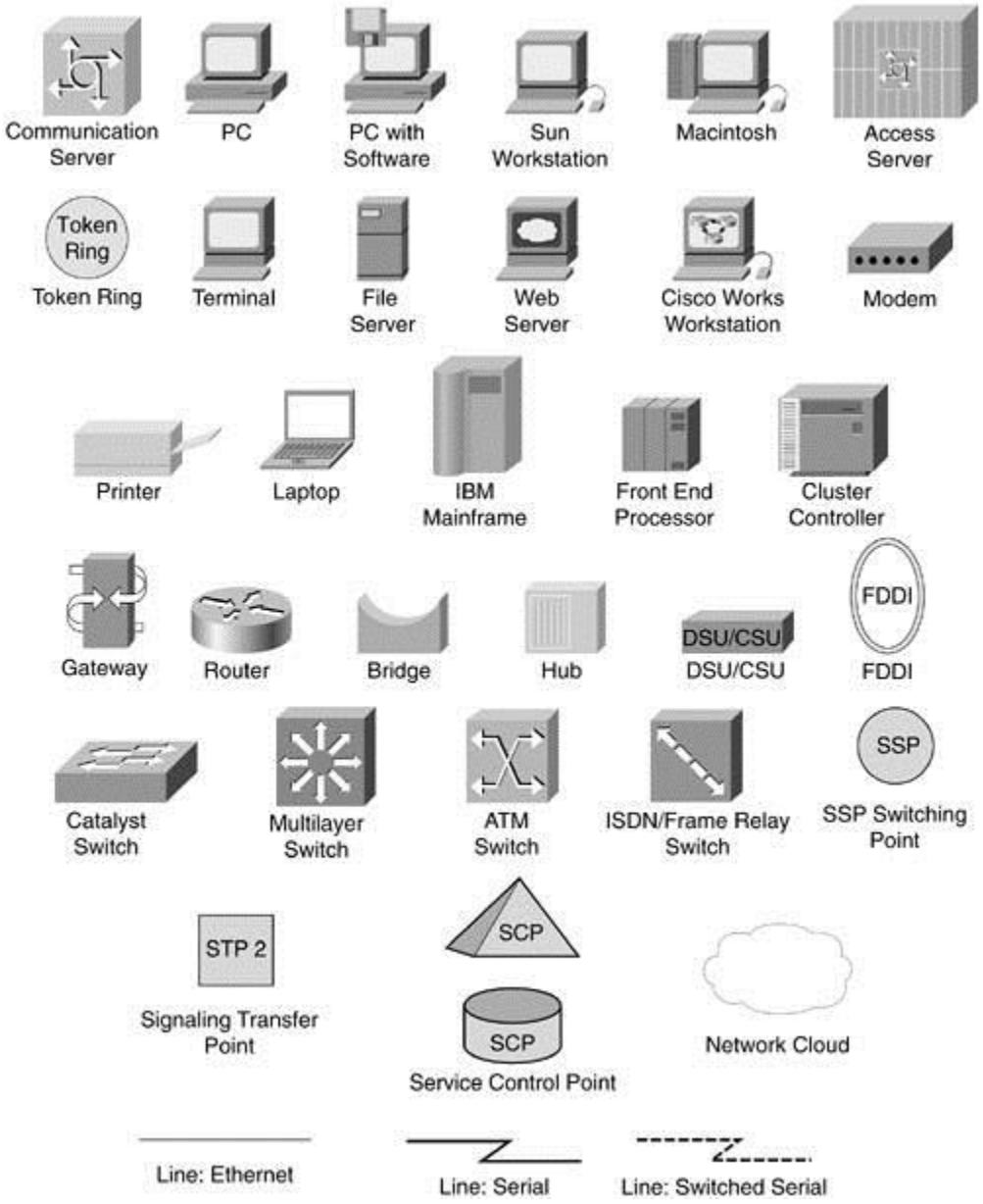
Signaling System No. 7 (SS7/C7): Protocol, Architecture, and Services

By Lee Dryburgh, Jeff Hewett

Publisher: Cisco Press
Pub Date: August 02, 2004
ISBN: 1-58705-040-4
Pages: 744

- Table of Contents
- Index

Icons Used in This Book



Command Syntax Conventions

The conventions used to present command syntax in this book are the same as the conventions used in the IOS Command Reference. The Command Reference describes these conventions as follows:

- Vertical bars (|) separate alternative, mutually exclusive elements.

- Square brackets ([ ]) indicate an optional element.
- Braces ({ }) indicate a required choice.
- Braces within brackets ([{ }]) indicate a required choice within an optional element.
- **Bold** indicates commands and keywords that are entered literally as shown. In actual configuration examples and output (not general command syntax), bold indicates commands that are manually input by the user (such as a **show** command).
- *Italic* indicates arguments for which you supply actual values.

Introduction

SS7/C7 is a signaling network and protocol that is used worldwide to bring telecommunications networks, both fixed-line and cellular, to life. Setting up phone calls, providing cellular roaming and messaging, and converged voice/data services, such as Internet Call Waiting, are only a few of the vast number of ways that SS7/C7 is used in the communications network. SS7/C7 is at the very heart of telecommunications, and as voice networks and data networks converge, SS7/C7 will provide the technology to bridge the gap between the two worlds. Anyone who is interested in telecommunications should have a solid understanding of SS7/C7. The convergence of voice and data has extended the need to understand this technology into the realm of those working with data networks.

How This Book Is Organized

Those who are new to the world of telecommunications signaling should read Chapters 1 to 5 first, in sequence. Those who are already comfortable with telecommunications and signaling concepts can read particular chapters of interest. This book should prove the most valuable for those who already consider themselves experts in SS7/C7; in particular, attention should be given to the extensive appendixes.

Part I: Introductions and Overviews

Chapter 1: The Evolution of Signaling

This chapter introduces the concept of signaling. It is a great starting point for those who are unfamiliar with signaling or telecommunications in general. It introduces concepts and terminology that are used throughout the book.

Chapter 2: Standards

This chapter introduces the relevant standards and the bodies that are involved in creating them. It also provides some background on both the history of the standards and the bodies themselves. In addition, it introduces the concept of standards on different planes—national, regional, and international.

Chapter 3: The Role of SS7

This chapter is an excellent introduction to SS7/C7 and its relevance. Any reader can read it, regardless of background. Hopefully even those who are very knowledgeable in SS7/C7 will find this chapter interesting, because it lists the functions and services offered by SS7/C7 and explains its relevance in the daily lives of people across the globe.

Chapter 4: SS7 Network Architecture and Protocols Introduction

This chapter provides a technical overview of the SS7 protocol and network architecture. Those who are new to the subject will find it particularly interesting. It provides an introductory technical overview of SS7 in such a way that newcomers can assimilate subsequent chapters more effectively.

Chapter 5: The Public Switched Telephone Network (PSTN)

This chapter provides a brief overview of the Public Switched Telephone Network. It helps you understand SS7 in its native environment as the primary form of interoffice signaling in the PSTN. It also briefly introduces the PSTN's transition to the next-generation Voice Over Packet architecture.

Part II: Protocols Found in the Traditional SS7/C7 Stack

Chapter 6: Message Transfer Part 2 (MTP2)

This chapter examines the first protocol on top of the physical layer. It covers frame format, functions, and procedures—packet delineation, error correction, error detection, alignment, managing the signaling link, procedures for establishing a signaling link, flow control, and link error monitoring.

Chapter 7: Message Transfer Part 3 (MTP3)

This chapter covers the core concepts of how SS7 network nodes communicate with each other. It discusses network addressing and routing in detail, along with examples of how messages flow through an SS7 node. It also explains the numerous messages and procedures that MTP3 uses to maintain a healthy network.

Chapter 8: ISDN User Part (ISUP)

This chapter explains how the ISUP portion of the protocol is used to set up and tear down calls, provide trunk maintenance functions, and deliver supplementary services. It defines ISUP message structure as well as the most commonly used messages and parameters. The association between call processing at an SSP and the ISUP protocol is described, thereby helping you understand how an SS7-signaled call is processed at an SSP.

Chapter 9: Signaling Connection Control Part (SCCP)

This chapter looks at the enhanced functionality that the addition of this protocol brings—namely, application management, more flexible and powerful routing through the use of global titles, and mechanisms for transferring application data over the signaling network.

Chapter 10: Transaction Capabilities Application Part (TCAP)

This chapter describes the role of TCAP in providing a generic protocol mechanism for transferring information components between applications across the network. It helps you understand the key role TCAP plays in communication between SSP and SCP nodes. TCAP message formats and component definitions, including ITU and ANSI formats, are explained.

Part III: Service-Oriented Protocols

Chapter 11: Intelligent Networks (IN)

This chapter explains the concept of the Intelligent Network, how it has evolved, and how it is used to implement telecommunications services. It provides a detailed explanation of the IN call model and explains the parallels and differences between the ITU CS model and the North American AIN model. Several examples of IN services, such as toll-free calling and local number portability, are included to show how IN services are used.

Chapter 12: Cellular Networks

This chapter introduces GSM public land mobile networks (PLMNs) so that the following chapter can cover additional SS7 protocols used in cellular networks. It introduces cellular network entities, addressing, terminology, and concepts.

Chapter 13: GSM and ANSI-41 Mobile Application Part (MAP)

This chapter explains the operations and associated procedures that allow cellular subscribers to have mobility; this is the key functionality expected of a cellular network. Subscriber authentication, operations and maintenance, supplementary service, unstructured supplementary service (USS), and short message service (SMS) operations and procedures are also detailed.

Part IV: SS7/C7 Over IP

Chapter 14: SS7 in the Converged World

This chapter introduces the next-generation network architecture using media gateway controllers, media gateways, and signaling gateways. Its primary purpose is to provide an in-depth look at the Signaling Transport protocol (Sigtran), used between the media gateway controller and the signaling gateway. Sigtran is particularly interesting to those who are learning about SS7, because it provides a common signaling protocol interface between legacy SS7 networks and voice over IP networks.

Part V: Supplementary Topics

Chapter 15: SS7 Security and Monitoring

This chapter explains the need for SS7/C7 security practices. It describes the current means of providing security: traffic screening and monitoring. Details of providing traffic screening are supplied. Monitoring functionality and what should be monitored also are covered.

Chapter 16: SS7 Testing

This chapter explains the tools used for SS7/C7 protocol verification and how to create appropriate test specifications. It also outlines sample test cases for each protocol layer.

Part VI: Appendixes

Appendix A: MTP Messages (ANSI/ETSI/ITU)

This appendix lists all of the messages used by MTP3 for ANSI- and ITU-based networks. It also lists the message codes.

Appendix B: ISUP Messages (ANSI/UK/ETSI/ITU-T)

This appendix lists all of the messages used by ISUP for ANSI- and ITU-based networks. It also lists the message codes.

Appendix C: SCCP Messages (ANSI/ETSI/ITU-T)

This appendix lists all of the messages used by SCCP for ANSI- and ITU-based networks. It also lists the message codes.

Appendix D: TCAP Messages and Components

This appendix lists all of the messages and components used by MTP3 for ANSI- and ITU-based networks. It also lists the message codes.

Appendix E: ITU-T Q.931 Messages

Q.931 is the Layer 3 protocol of the subscriber signaling system that is used for ISDN, known as Digital Subscriber Signaling System No. 1 (DSS 1). It employs a message set that is made for interworking with SS7's ISUP. This appendix lists and describes the purpose of the Q.931 message set.

Appendix F: GSM and ANSI MAP Operations

This appendix lists the operations found in GSM MAP and their respective codes.

Appendix G: MTP Timers in ITU-T/ETSI/ANSI Applications

This appendix lists ANSI- and ITU-specified MTP timers.

Appendix H: ISUP Timers for ANSI/ETSI/ITU-T Applications

This appendix lists ANSI- and ITU-specified ISUP timers.

Appendix I: GSM Mobile Country Codes (MCC) and Mobile Network Codes (MNC)

This appendix lists all of the MCC codes and the respective MNCs found against the MCC.

Appendix J: ITU and ANSI Protocol Comparison

This appendix covers some of the main differences between ANSI and ITU (international).

Appendix K: SS7 Standards

This appendix presents the main SS7 standards alongside the respective standards body.

Appendix L: Tektronix Supporting Traffic

This appendix contains reference traffic caught on a Tektronix K1297 protocol analyzer.

Appendix M: Cause Values

Cause values, which are included as a field in each ISUP REL message, indicate why a call was released. This appendix lists and defines the ITU-T and ANSI cause values.

Chapter 1. The Evolution of Signaling

This chapter is intended to provide a sound introduction to the world of telecommunications signaling. It is particularly written for those readers who have little or no signaling knowledge. It provides a solid foundation to help you grasp signaling ideas, concepts, terminology, and methods. A strong foundation will provide the novice reader with a better understanding of the book's main topic: Signaling System No. 7. Today, Signaling System No. 7 is the most advanced and widely used signaling system for both cellular and fixed-line telecommunications networks.

This chapter covers the following topics:

- What signaling is and why it is relevant
- Overview of subscriber and network signaling
- The history of signaling and the development of the Public Switched Telephone Network (PSTN)
- Overview of the Channel Associated Signaling (CAS) method of signaling and its common implementations
- Overview of the Common Channel Signaling (CCS) method of signaling and its operational modes
- The limitations of CAS and CCS

Signaling System No. 7, known more commonly in North America as SS7 and elsewhere as C7, is both a network architecture and a series of protocols that provide telecommunications signaling. In order to begin studying SS7, you must first learn what telecommunications signaling is by studying its origins and purpose.

The ITU-T defines signaling as, [47] "The exchange of information (other than by speech) specifically concerned with the establishment, release and other control of calls, and network management, in automatic telecommunications operation."

In telecommunications, the network's components must indicate (that is, signal) certain information to each other to coordinate themselves for providing services. As such, the signaling network can be considered the telecommunications network's nervous system. It

breathes life into the infrastructure. Richard Manterfield, author of *Telecommunications Signaling*, has stated this poetically [103]:

"Without signaling, networks would be inert and passive aggregates of components. Signaling is the bond that provides dynamism and animation, transforming inert components into a living, cohesive and powerful medium."

For example, if a subscriber wishes to place a call, the call must be signaled to the subscriber's local switch. The initial signal in this process is the off-hook condition the subscriber causes by lifting the handset. The action of lifting the handset signals to the network that the subscriber wishes to engage telephony services. The local switch should then acknowledge the request for telephony services by sending back a dial tone, which informs the subscriber that he can proceed to dial the called party number. The subscriber has a certain amount of time to respond to the dial tone by using the telephone keypad to signal the digits that comprise the called party number. The network signals that it is receiving the dialed digits with silence (as opposed to a dial tone).

Up to this point, the signaling is known as *subscriber signaling* and takes place between the subscriber and the local switch. Subscriber signaling is also known as *access signaling*. The "Subscriber Signaling" section of this chapter further describes subscriber signaling.

NOTE

The calling party is often referred to as the A party. Similarly, the called party is referred to as the B party.

When a complete called party number is received or enough digits are collected to allow the routing process to proceed, the calling party's local switch begins signaling to the other nodes that form part of the core network.

The signaling that takes place between core network nodes (and switches and, over the past two decades, databases) is known as network signaling.

NOTE

Switches are also known as exchanges; within the United States, the term exchange is used interchangeably with Central Office (CO) or End Office (EO).

Network signaling is also known as inter-switch signaling, network-network signaling, or trunk signaling.

The purpose of network signaling is to set up a circuit between the calling and called parties so that user traffic (voice, fax, and analog dial-up modem, for example) can be transported bi-directionally. When a circuit is reserved between both parties, the destination local switch places a ringing signal to alert the called party about the incoming call. This signal is classified as subscriber signaling because it travels between a switch (the called party's local switch) and a subscriber (the called party). A ringing indication tone is sent to the ~~calling~~ called party telephone to signal that the telephone is ringing. If the called party wishes to engage the call, the subscriber lifts the handset into the off-hook condition. This moves the call from the set-up phase to the call phase.

At some point in the call phase, one of the parties will wish to terminate the call, thereby ending the call phase. The calling party typically initiates this final phase, which is known as the clear-down or release phase. The subscriber signals the network of the wish to terminate a call by placing the telephone back in the on-hook condition; hence, subscriber signaling. The local switch proceeds with network signaling to clear the call down. This places an expensive resource (the circuit) back to an idle condition, where it can be reserved for another call.

The previous high-level example relates to a basic telephone service call; that is, simple call setup and clear down. As you will discover, the signaling network can do far more than carry the digits you dial, release calls, notify the network that you went on or off-hook, and so forth. The signaling network can also translate toll-free numbers into "routable" numbers, validate credit and calling cards, provide billing information, remove faulty trunks from service, provide the support for supplementary services (such as caller ID), allow you to roam with your cellular telephone, and makes *local number portability (LNP)* possible. This list is by no means exhaustive; see Chapters 3, "The Role of SS7," and 11, "Intelligent Networks (IN)," for more example services.

The main function of signaling is still that of circuit supervision: setting up and clearing down circuits (that is, trunks). Traditionally, once a circuit was set up, no other signaling was performed apart from releasing the call; therefore, all calls were simple, basic telephone service calls. However, modern telephone networks can perform signaling while a call is in progress, especially for supplementary services—for example, to introduce another called party into the call, or to signal the arrival of another incoming call (call waiting) to one of the parties. In fact, since the 1980s, signaling can take place even when there is not a call in place. This is known as *non-circuit related signaling* and is simply used to transfer data between networks nodes. It is primarily used for query and response with telecommunications databases to support cellular networks, intelligent networks, and supplementary services. For example, in Public Land Mobile Networks (PLMNs), the visitor location register (VLR) that is in charge of the area into which the subscriber has roamed updates the home location register (HLR) of the subscriber's location. PLMNs make much use of non-circuit-related signaling, particularly to keep track of roaming subscribers. Chapter 13, "GSM and ANSI-41 Mobile Application Part (MAP)," covers this topic in more detail.

Network signaling is further described in the "Network Signaling" section of this chapter.

The History of Signaling

To appreciate signaling in today's network and its role in future networks, let's examine the history of signaling. The history of signaling has been inextricably linked to the history of telecommunications and, in particular, switching. As telecommunications advances, so do the signaling systems that support it.

1889–1976

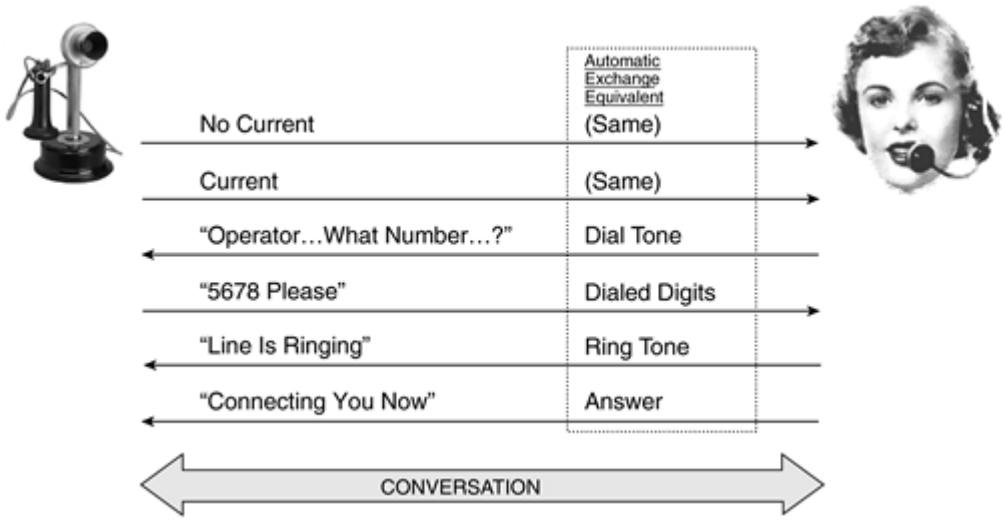
The earliest telephone switches were manual; operators used a switchboard and wire cords to connect and disconnect all calls. The first manual exchange occurred in 1878 in New

Haven, Connecticut. It was introduced to avoid the imminent problem of running wires from each telephone to every other telephone (a fully meshed topology). The first manual switch appeared in Great Britain in 1879. It was also within this same year that subscribers came to be called by numbers rather than by names. Within a decade of introducing the manual switch, the United States had 140,000 subscribers and a staggering 8000 exchanges—that is, a switch for every 17.5 subscribers!

A subscriber who was connected to a manual switch would crank a lever to electronically send an alerting signal that lit up a bulb on the operator's switchboard. The operator would then connect her telephone to the calling line, and ask for the called number. Next the operator would connect her telephone to the called line, where she would place a ringing signal. If the called party answered the call, the operator would establish the connection by plugging in a cord between the two terminal jacks on the switchboard. Figure 1-1 shows this process; on the switchboard, each terminal jack represents a subscriber.

Figure 1-1. Simple Call Setup Via a Manual Operator with Automatic Equivalent

[View full size image]



Signaling, as we know it today, began around 1889 with the invention of the Strowger exchange (which was patented 1891). The Strowger exchange was an electromechanical device that provided automatic switching using the simple idea of two-motion selectors for establishing calls between two subscribers. It was also known as a step-by-step switch because it followed pre-wired switching stages from start to finish.

Inventing the Strowger Exchange

Almon B. Strowger was a schoolteacher and part-time undertaker. His reportedly constant feuds with manual switchboard operators inspired him to develop an automatic switching

system and the dial telephone so he could bypass manual switchboard operators [102]. One reported feud concerned an alleged business loss resulting from the complete lack of privacy offered by a manual exchange. Strowger claimed that an operator at the new manual exchange in Connecticut had intentionally directed a call to a competitor—an allegation that gave rise to tales that the operator was either married to or was the daughter of a competing undertaker. Strowger moved from Topeka to Kansas City, where he hoped his new, larger funeral home would earn him his fortune. However, he suffered a similar fate there; he believed that the manual operators there were intentionally giving his customers a busy signal. Strowger therefore decided to do away with operators; he hired several electromechanical technicians, who created the first automatic exchange within a year. As a result, the telephone became faster, easier to use, and more private for everyone.

The first Strowger exchange in the United States opened in La Porte, Indiana in 1892 and had the switching capacity for ninety-nine lines. Lobby groups protested at the automatic exchange, and one lobby group championed the personalized service afforded by manual exchanges. The lobby group did not have much success, however; manual switchboards could not service the dramatic increase in telephone subscribers. By 1900 there were 1.4 million telephones in the United States.

In Great Britain, the first Strowger exchange opened at Epsom in Surrey in 1912. The last Strowger switch was not removed from the British Telecom (BT) service network until June 23, 1995, when it was removed from Crawford, Scotland.

Strowger sold his patents to his associates for $1,800 in 1896 and sold his share in the company for $10,000 in 1898. He died in 1902. In 1916, his patents were sold to Bell Systems for $2.5 million dollars.

Strowgers' dial telephone is considered the precursor of today's touch-tone phone. It had three buttons: one for hundreds, one for tens, and one for units. To call the number 322, the caller had to push the hundreds button three times, the tens button two times, and the units button two times.

In 1896 the Automatic Electric Company developed a rotary dial to generate the pulses. This method of transmitting the dialed digits became known as *pulse dialing* and was commonplace until the latter half of the twentieth century, when *tone dialing* became available. See "Address Signals" in the "Subscriber Signaling" section of this chapter for a discussion of pulse and touch-tone dialing. It is interesting to note that early users did not like the dial pulse handset because they felt they were doing the "telephone company's job."

Even in Great Britain in 1930, the majority of all local and long distance calls were still connected manually through an operator. But gradually, calls placed between subscribers served by the same local switch could be dialed without the help of an operator. Therefore, only subscriber signaling was required because an operator would perform any inter-switch signaling manually. In the decades that followed, it became possible to dial calls between subscribers who were served by nearby switches. Thus the requirement for network signaling was born. Most large U.S. cities had automatic exchanges by 1940.

Direct Distance Dialing (*DDD*) was introduced in the United States in the 1950s. DDD allowed national long distance calls to be placed without operator assistance, meaning that

any switch in the United States could route signaling to any other switch in the country. *International Direct Distance Dialing* (IDDD) became possible in the 1960s, thus creating the requirement for signaling between international switches.

From 1889 to 1976, signaling had three main characteristics, which resulted because only basic telephone services were available [102]:

- Signaling was fairly simple. All that was required of the signaling system was the setting-up and releasing of circuits between two subscribers.
- Signaling was always circuit-related; that is, all signals related directly to the setting-up or clearing of circuits.
- There was a deterministic relationship, known as *Channel Associated Signaling (CAS)*, between the signaling and the voice traffic it controlled. The "Channel Associated Signaling" section of this chapter discusses CAS.

1976 to Present Day

Another form of signaling was introduced in 1976: *Common Channel Signaling* (CCS). The "Common Channel Signaling" section of this chapter further explains CSS.

CCS has been used to implement applications beyond the scope of basic telephone service, including Intelligent Networks (INs), supplementary services, and signaling in cellular mobile networks. As you will learn, SS7 is the modern day CCS system that is used for network signaling. As with any technical subject, signaling can be split into a number of classifications. The broadest classification is whether the signaling is subscriber or networked signaling. The following sections discuss these types of signaling

Subscriber Signaling

Subscriber signaling takes place on the line between the subscribers and their local switch. Most subscribers are connected to their local switch by analog subscriber lines as opposed to a digital connection provided by an Integrated Services Digital Network (ISDN). As a result, subscriber signaling has evolved less rapidly than network signaling.

Subscriber signals can be broken down into the following four categories:

- Address Signals
- Supervisory Signals
- Tones and Announcements
- Ringing

Address Signals

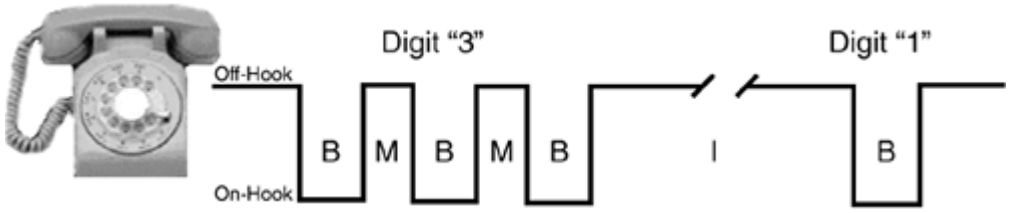
Address signals represent the called party number's dialed digits. Address signaling occurs when the telephone is off-hook. For analog lines, address signaling is either conveyed by the dial pulse or Dual-Tone Multiple Frequency (DTMF) methods. Local switches can typically handle both types of address signaling, but the vast majority of subscribers now use Dual-Tone Multi Frequency (DTMF), also known as touch-tone.

The precursor to (DTMF) was dial pulse, which is also known as rotary dialing. In rotary dialing, the address signals are generated by a dial that interrupts the steady DC current at a sequence determined by the selected digit. The dial is rotated clockwise, according to the digit selected by the user. A spring is wound as the dial is turned; when the dial is subsequently released, the spring causes the dial to rotate back to its original resting position. Inside the dial, a governor device ensures a constant rate of return rotation, and a shaft on the governor turns a cam that opens and closes switch contact. The current flowing into the telephone handset is stopped when the switch contact is open, thereby creating a dial pulse. As the dial rotates, it opens and closes an electrical circuit.

The number of breaks in the string represents the digits: one break for value 1, two breaks for value 2, and so on (except for the value of 0, which is signaled using ten breaks). The nominal value for a break is 60 ms. The breaks are spaced with make intervals of nominally 40 ms. As shown in Figure 1-2, consecutive digits are separated by an inter-digit interval of a value greater than 300 ms.

Figure 1-2. Dial Pulse Address Signals

[View full size image]



The rotary dial was designed for operating an electromechanical switching system; the speed of the dial's operation was approximately to match the switches' operating speed.

DTMF is a modern improvement on pulse dialing that first appeared during the 1960s and is now widespread. A DTMF signal is created using a pair of tones, each with a different frequency. It is much faster than the previous pulse method and can be used for signaling after call completion (for example, to operate electronic menu systems or activate supplementary services, such as a three-way call). The standard DTMF has two more buttons than dial pulse systems: the star (*) and the pound, or hash (#) buttons. These buttons are typically used in data services and customer-controlled features. The CCITT has standardized the DTMF frequency combinations, as shown in Table 1-1. For additional information regarding the CCITT, see Chapter 2, "Standards."

Table 1-1. Tones Used to Create DTMF Signals

	1209 Hz	**1336 Hz**	**1477 Hz**	**1633 Hz**
697 Hz	1	2	3	A
770 Hz	4	5	6	B
852 Hz	7	8	9	C
941 Hz	*	0	#	D

The fourth column (1633 Hz) has several special uses that are not found on regular telephones. The four extra digits were used on special handsets to designate the priority of calls on the Automatic Voice Network (AUTOVON), the U.S. military phone network that has since been replaced with the Defense Switched Network (DSN). In AUTOVON, the keys were called Flash, Immediate, Priority, and Routine (with variations) instead of ABCD. Telephone companies still use the extra keys on test handsets for specific testing purposes.

All modern telephone handsets support both DTMF and dial pulse. Because an electronic handset has buttons rather than a rotary dial, the numbers are temporally stored in the telephone memory to generate pulse dialing. The handset then transmits the dial pulses. This arrangement is sometimes known as *digipulse*.

Supervisory Signals

A telephone has two possible supervision states: *on-hook* and ~~or~~ *off-hook*. On-hook is the condition in which the telephone is not in use, which is signaled when the telephone handset depresses the cradle switch. The term on-hook comes from the days when the receiver part of the telephone rested on a hook. The telephone enters the off-hook condition when the handset is lifted from its cradle, thereby releasing the cradle switch and signaling to the exchange that the subscriber wishes to place an outgoing call.

Residential systems worldwide use a change in electrical conditions, known as *loop start signaling*, to indicate supervision signals. The local switch provides a nominal –48 V direct current (DC) battery, which has the potential to flow through the subscriber line (between the local switch and the subscriber). When a telephone is off-hook, DC can flow in the subscriber line; when a telephone is on-hook a capacitor blocks the DC. The presence or absence of direct current in the subscriber's local switch line determines the telephone's supervision state. Loop start systems are adequate for residential use, but a problem known as *glare* makes loop start unacceptable in typical business applications in which private exchanges (PBXs) are used. PBXs use a system known as *ground start signaling*, particularly in North America.

Ground start systems combat glare by allowing the network to indicate off-hook (seizure) for incoming calls, regardless of the ringing signal. This reduces the probability of simultaneous seizure, or glare, from both ends. Ground start requires both ground and current detectors in customer premise equipment (CPE).

Tones and Announcements

Tones and announcements are audible backward signals, such as dial tone, ring back, and busy-tone, that are sent by a switch to the calling party to indicate a call's progress. Table 1-2 shows the call progress tones that are used in North America.

Table 1-2. Call Progress Tones Used in North America

Tone	Frequency (Hz)	On Time (Sec)	Off Time (Sec)
Dial	350+440	Continuous	
Busy	480+620	0.5	0.5
Ring back, Normal	440+480	2	4
Ring back, PBX	440+488	1	3
Congestion (Local)	480+620	0.3	0.2
Congestion (Toll)	480+620	0.2	0.3
Howler (Receiver wrongly off-hook)	1400+2060+2450+2600	0.1	0.1

Forward signals refer to signals that transfer in the direction of call establishment, or from the calling party to the called party. Backward signals refer to signals that transfer in the reverse direction.

Ringing

Ringing is a forward signal sent by the switch to the called subscriber to indicate the arrival of a call. It is known more specifically as power ringing to distinguish it from audible ringing, which is played to the calling party to alert him that the called party phone is ringing. Each country has a ringing pattern, which is known as the *cadence*. In North America the pattern is two seconds on, four seconds off.

Note that audible and power ringing are not synchronized. This is why, on a rare occasion, a caller is already on the line when you lift the handset. This situation generally causes confusion because the calling party, who has heard audible ringing, is unaware of the problem since the problem occurs because the caller's switch does not generate an independent ringing signal for each line. Instead, it generates one signal that is applied to whichever lines are to be played audible ringing. Therefore, if you have an incoming call, the switch must wait until the next on-cycle to ring your telephone. If you happen to pick up the telephone during the few off-cycle seconds and a call has just come in, you have answered a call before the exchange has had the opportunity to alert you of the incoming call. In North America, the silent period during which inbound calls cannot be announced is 3.9 seconds. Countries that use a short period of silence in the ringing cadence are less susceptible to this problem.

NOTE

If you are one of those people who say that you will call home and let the telephone ring twice when you get to your destination safely, note that you have no guarantee that the telephone will actually ring twice—or even ring at all. You might hear two rings, but that does not mean the called party will hear two, or even any, rings because their power ringing pattern might be in an off period.

The problems associated with the lack of synchronization between the calling and called party is typically addressed in North American non-residential systems (PBX systems) by using ground start rather than loop start. Other countries often employ a simple technique known as *ring splash*. With ring splash, a PBX issues a brief ringing tone within a few hundred milliseconds of the trunk being seized (the incoming call), after which normal ringing cadence resumes. The downside to this solution is that the ringing cadence sounds strange because it is not synchronized with the initial ring.

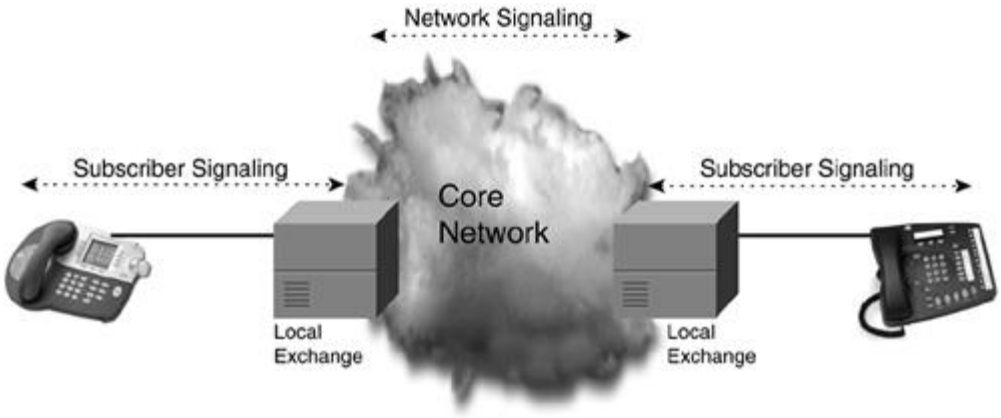
Network Signaling

As previously described, network signaling takes place between nodes in the core network. This is generally from the local switch, through the core network, and to the destination local switch—in other words, between the calling and the called party switch.

Figure 1-3 shows where subscriber and network signaling occur in the PSTN.

Figure 1-3. Subscriber and Network Signaling

[View full size image]



For obvious reasons, the signaling system employed on the local loop (between the subscriber and the local switch) differs from that which is used in the core network. The subscriber must only generate a limited number of signals: on or off hook, called party digits, and possibly a few commands for supplementary services. In comparison, a modern core network must perform very complex signaling, such as those to support database driven services like Local Number Portability (LNP), credit or calling card validation, and cellular roaming. Therefore, subscriber signaling systems are simple compared to modern network signaling systems.

Network signaling was previously implemented using *Channel Associated Signaling (CAS)* techniques and systems. However, for the past two decades, it has been replaced with *Common Channel Signaling (CCS)* systems. Apart from a rare trace of Signaling System No. 6 (SS6) signaling, System No. 7 (SS7) is almost the exclusive CSS system; thus, CCS can almost be taken to refer exclusively to the use of SS7. The remaining sections of this chapter discuss CAS and CCS methods.

Channel Associated Signaling

The key feature that distinguishes Channel Associated Signaling (CAS) from CCS is the deterministic relationship between the call-control signals and the bearers (voice circuits) they control in CAS systems. In other words, a dedicated fixed signaling capacity is set aside for each and every trunk in a fixed, pre-determined way.

Channel Associated Signaling (CAS) is often still used for international signaling; national systems in richer nations almost exclusively use Common Channel Signaling (CCS). CCS is replacing CAS on international interfaces.

CAS can be implemented using the following related systems:

- Bell Systems MF, R2, R1, and C5.
- Single-frequency (SF) in-band and out-of-band signaling
- Robbed bit signaling

The following sections discuss these methods in context with the type of signal, either address or supervisory.

Address Signals

Multifrequency systems, such as the Bell System MF, R2, R1, and C5, are all types of address signals used by CAS.

Multifrequency

The CAS system can be used on either analog Frequency Division Multiplexed (FDM) or digital Time Division Multiplexed (TDM) trunks. MF is used to signal the address digits between the switches.

Multifrequency (MF) signaling can still be found in traces within the United States, and it is still often found on international interfaces. On international interfaces outside of North America, MF is still used via the CCITT System 5 (C5) implementation. C5 is quite similar to Bell MF and was developed jointly by Bell Laboratories and the British Post Office [102]. R2 is the MF system that was deployed outside North America and is still used in less developed nations. R2 was developed by CEPT (which later became ETSI; see Chapter 2) and was previously known as Multifrequency Compelled (MFC) signaling. The CCITT later defined an international version; see Chapter 2 for additional information regarding the international version [102].

MF simultaneously sends two frequencies, from a choice of six, to convey an address signal. The switch indicates to the switch on the other end of a trunk that it wishes to transmit

address digits by sending the KP (start pulsing) signal, and indicates the end of address digits by sending the ST (end pulsing) signal. The timing of MF signals is a nominal 60 ms, except for KP, which has a nominal duration of 100 ms. A nominal 60 ms should be between digits.

Table 1-3 shows the tone combinations for Bell System MF, R1, and C5. R2 tone combinations are not shown.

Table 1-3. Tones Used to Create MF Signals

Digit	Frequencies					
	700	900	1100	1300	1500	1700
1	+	+				
2	+		+			
3		+	+			
4	+			+		
5		+		+		
6			+	+		
7	+				+	
8		+			+	
9			+		+	
0				+	+	
KP			+			+
ST					+	+
11 [*]	+					+
12 [*]		+				+

Table 1-3. Tones Used to Create MF Signals

Digit	Frequencies					
	700	900	1100	1300	1500	1700
KP2 [*]				+		+

[*] = Used only on CCITT System 5 (C5) for international calling.

As stated, many international trunks still use C5. Signal KP2 indicates that the number is an international number; by inference, KP indicates that the number is a national number. International operators also use codes 11 and 12. More details on C5 are available in ITU-T Q.152. Supervision signals for MF systems are performed on FDM trunks by the use of Single Frequency (SF), which we describe in the following section.

For circuit supervision, both Bell System MF and R1 use Single Frequency (SF) on FDM trunks and employ robbed bit signaling on TDM controlled trunks. C5 uses a different set of MF tones for supervisory signaling.

Supervisory Signals

Single frequency systems, robbed bit signaling, and digital signaling are all types of *supervisory signals* used by CAS.

Single Frequency(SF)

Single Frequency (SF) was used for supervisory signaling in analog CAS-based systems. North America used a frequency of 2600 Hz (1600 Hz was previously used), and Great Britain used 2280 Hz (as defined in British Telecom's SSAC15 signaling specification). When in an on-hook state, the tone is present; when in an off-hook state, the tone is dropped.

NOTE

Supervisory signals operate similarly to those used in access signaling; however, they signal the trunk state between two switches rather than the intention to place or terminate a call. Supervisory signals are also known as line signals.

Table 1-4 details the tone transitions Bell System MF and R1 use to indicate the supervision signals. C5 uses a combination of both one and two in-band signaling tones, which are not presented here.

Table 1-4. Bell System MF and R1 Supervision Signaling

Direction	Signal Type	Transition

Table 1-4. Bell System MF and R1 Supervision Signaling

Direction	Signal Type	Transition
Forward	Seizure	On-hook to off-hook
Forward	Clear-forward	Off-hook to on-hook
Backward	Answer	On-hook to off-hook
Backward	Clear-back	Off-hook to on-hook
Backward	Proceed-to-send (wink)	Off-hook pulse, 120–290 ms

As with the MF address signaling, SF is sent switch to switch. A trunk is initially on-hook at both ends. One of the switches sends a forward off-hook (seizure) to reserve a trunk. The receiving switch indicates that it is ready to receive address digits, (after connecting a digit received by the line by sending a wink signal. When the originating switch receives the wink signal, it transmits the digits of the called party number. When a call is answered, the called parties switch sends an off-hook signal (answer). During the conversation phase, both ends at each trunk are off-hook. If the calling a party clears the call, it sends a clear-forward signal; likewise, when the called party hangs up, it sends a clear-backward signal.

SF uses an *in-band* tone. In-band systems send the signaling information within the user's voice frequency range (300 Hz to 3400 Hz). A major problem with in-band supervisory signaling, however, is its susceptibility to fraud. The hacker quarterly magazine "2600" was named for the infamous 2600 Hz tone, which could be used by the public to trick the phone system into giving out free calls. The subscriber could send supervisory tone sequences down his telephone's mouthpiece using a handheld tone generator. This enabled the subscriber to instruct switches and, in doing so, illegally place free telephone calls.

The other major problem with in-band signaling is its contention with user traffic (speech). Because they share the same frequency bandwidth, only signaling or user traffic can be present at any one time. Therefore, in-band signaling is restricted to setting up and clearing calls down only because signaling is not possible once a call is in progress.

Subscriber Line Signaling

A regular subscriber line (that is analog) still uses in-band access signaling. For example, DTMF is used to signal the dialed digits and the frequencies used are within the voice band (see Table 1-1). You can prove that DTMF uses in-band signaling by using a device, such as a computer, to generate the tones for each digit (with correct pauses). Simply play the tones from the computer speaker down the mouthpiece of a touch-tone telephone. This allows you to dial a number without using the telephone keypad. Because the signaling is sent down the mouthpiece, you can be certain that it traveled within the user's voice frequency range.

FDM analog systems nearly always reserve up to 4000 Hz for each circuit, but only use 300–3400 Hz for speech; therefore, signaling is sent above the 3400 Hz (and below 4000 Hz). This is known as out-of-band signaling and is used in R2 for supervisory signaling. Unlike with in-band signaling, no contention exists between user traffic and signaling. North America uses a frequency of 3700 Hz, and CCITT (international) uses 3825 Hz. Table 1-5 details the tone transitions that indicate the supervision signals used in R2 and R1.

Table 1-5. R2 Supervision Signaling

Direction	Signal Type	Transition
Forward	Seizure	Tone-on to tone-off
Forward	Clear-forward	Tone-off to tone-on
Backward	Answer	Tone-on to tone-off
Backward	Clear-back	Tone-off to tone-on
Backward	Release-guard	450 ms tone-off pulse
Backward	Blocking	Tone-on to tone-off

R2 does not use a proceed-to-send signal; instead, it includes a blocking signal to stop the circuit that is being seized while maintenance work is performed on the trunk. The release guard signal indicates that the trunk has been released after a clear-forward signaling, thereby indicating that the trunk can be used for another call.

Digital

Supervisory signaling can be performed for R2 on digital TDM trunks. On an E1 facility, timeslot 16 is set aside for supervisory signaling bits (TS16). These bits are arranged in a multiframe structure so that specific bits in the multiframe's specific frames represent the signaling information for a given TDM audio channel. See Chapter 5, "The Public Switched Telephone Network (PSTN)," for explanation of facilities and timeslots.

Limitations of CAS

We discuss the general disadvantages of CAS for the purpose of reinforcing the concepts and principles we have introduced thus far. CAS has a number of limitations, including:

- Susceptibility to fraud
- Limited signaling states
- Poor resource usage/allocation

The following sections discuss these limitations in more detail.

Susceptibility to Fraud

CAS employing in-band supervisory signaling is extremely susceptible to fraud because the subscriber can generate these signals by simply using a tone generator down a handset

mouthpiece. This type of device is known as a *blue box*; from the beginning of the 1970s, it could be purchased as a small, handheld keypad. Blue box software was available for the personal computer by the beginning of the 1980s.

Limited Signaling Information

CAS is limited by the amount of information that can be signaled using the voice channel. Because only a small portion of the voice band is used for signaling, often CAS cannot meet the requirements of today's modern networks, which require much higher bandwidth signaling.

Inefficient Use of Resources

CAS systems are inefficient because they require either continuous signaling or, in the case of digital CAS, at regular intervals even without new signals.

In addition, there is contention between voice and signaling with in-band CAS. As a result, signaling is limited to call set-up and release phases only. This means that signaling cannot take place during the call connection phase, severely imposing technological limits on the system's complexity and usefulness.

Common Channel Signaling (CCS)

CCS refers to the situation in which the signaling capacity is provided in a common pool, with the capacity being used as and when necessary. The signaling channel can usually carry signaling information for thousands of traffic circuits.

In North America, signaling can be placed on its own T1 carrier even though it only takes up one timeslot. This means that two physical networks, "speech" and "signaling," can have different routings. (Please refer to Chapter 5 for a description of carriers and timeslots.) Alternatively, the signaling might exist on a carrier with other user traffic, depending on the network operator.

Outside of North America, the signaling is placed in its own timeslot on an E1 (that is, logically rather than physically separated). The other timeslots on E1 are for user traffic—apart from TS0, which is used for synchronization. E1 systems tend to use the TS16 timeslot for signaling; some core network equipment ignores TS16, expecting it to be used for signaling traffic because it has historically been the timeslot for digital CAS signaling.

The only CCS systems that have been implemented to date are Signaling Systems No. 6 and No. 7 (SS6 and SS7). The ITU for the international network originally standardized SS6, but they saw limited deployment. AT&T nationalized SS6 for the North American network and called it Common Channel Interoffice Signaling (CCIS) No. 6. SS6 saw a limited deployment after the mid-1970s because it had far less bandwidth and a much smaller packet size than SS7. In addition, its evolutionary potential was severely limited because it was not a layered protocol architecture.

CCS systems are packet-based, transferring over 200 bytes in a single SS7 packet, as opposed to a few bits allocated to act as indicators in digital CAS. The signaling information is transferred by means of *messages*, which is a block of information that is divided into

fields that define a certain parameter or further sub-field. The signaling system's specifications (Recommendations and Standards) define the structure of a message, including its fields and parameters.

Because CCS is packet-based and there is not a rigid tie between the signaling and the circuits it controls, it can operate in two distinct ways. These two distinct ways are circuit-related signaling and non-circuit-related signaling.

Circuit-Related Signaling

Circuit-related signaling refers to the original functionality of signaling, which is to establish, supervise, and release trunks. In other words, it is used to set up, manage, and clear down basic telephone service calls. Circuit-related signaling remains the most common mode of signaling. As it is with CAS, signaling capacity is not pre-allocated for each traffic circuit. Rather, it is allocated as it is required. Each signaling message is related to a traffic circuit. Because no dedicated relationship exists between the circuits and the signaling, it is necessary to identify the traffic circuit to which a particular signal message refers. This is achieved by including a circuit reference field in each signaling message.

like used in CAS

Non-Circuit-Related Signaling

Non-circuit-related signaling refers to signaling that is not related to the establishment, supervision, and release of trunks. Due to the advent of supplementary services and the need for database communication in cellular networks and Intelligent Networks, for example, signaling is no longer exclusively for simply setting up, managing, and clearing down traffic circuits. Non-circuit-related signaling allows the transfer of information that is not related to a particular circuit, typically for the purpose of transmitting both the query and response to and from telecommunication databases. Non-circuit-related signaling provides a means for transferring data freely between network entities without the constraint of being related to the control of traffic circuits.

Common Channel Signaling Modes

A signaling mode refers to the relationship between the traffic and the signaling path. Because CCS does not employ a fixed, deterministic relationship between the traffic circuits and the signaling, there is a great deal of scope for the two to have differing relationships to each other. These differing relationships are known as *signaling modes*.

There are three types of CCS signaling modes:

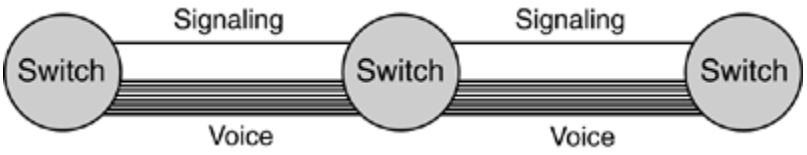
- Associated
- Quasi-associated
- Non-associated

SS7 runs in associated or quasi-associated mode, but not in non-associated mode. Associated and quasi-associated signaling modes ensure sequential delivery, while non-associated does not. SS7 does not run in non-associated mode because it does not have procedures for reordering out-of-sequence messages.

Associated Signaling

In *associated* mode, both the signaling and the corresponding user traffic take the same route through the network. Networks that employ only associated mode are easier to design and maintain; however, they are less economic, except in small-sized networks. Associated mode requires every network switch to have signaling links to every other interconnected switch (this is known as a fully meshed network design). Usually a minimum of two signaling links are employed for redundancy, even though the switched traffic between two interconnected switches might not justify such expensive provisioning. Associated signaling mode is the common means of implementation outside of North America. Figure 1-4 illustrates the associated concept.

Figure 1-4. Associated Mode

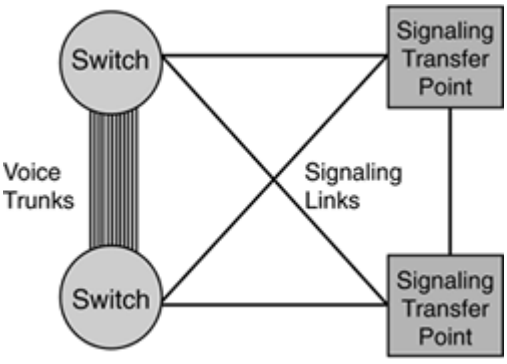


Quasi-Associated Signaling

In *quasi-associated* mode, signaling follows a different route than the switched traffic to which it refers, requiring the signaling to traverse at least one intermediate node. Quasi-associated networks tend to make better use of the signaling links; however, it also tends to create a more complex network in which failures have more potential to be catastrophic.

Quasi-associated signaling can be the most economical way of signaling for lightly loaded routes because it avoids the need for direct links. The signaling is routed through one or more intermediate nodes. Signaling packets arrive in sequence using quasi-associated signaling because the path is fixed for a given call (or database transaction) at the start of a call (or transaction). Figure 1-5 shows the quasi-associated signaling mode, which is the common means of implementation within North America.

Figure 1-5. Quasi-Associated Mode



Non-Associated Signaling

Because the path is not fixed at a given point in time in *non-associated* mode, the signaling has many possible routes through the network for a given call or transaction. Therefore, the packets might arrive out of sequence because different routes might have been traversed.

SS7 does not run in non-associated mode because no procedures exist for reordering out-of-sequence messages. Associated and quasi-associated signaling modes assure sequential delivery, while non-associated signaling does not. Quasi-associated mode is a limited case of non-associated mode, in which the relative path is fixed.

Summary

CCS has evolved to address the limitations of the CAS signaling method. CCS has the following advantages over CAS:

- Much faster call set-up time
- Greater flexibility
- Capacity to evolve
- More cost effective than CAS
- Greater call control

Most CCS calls can be set up in half the time it takes to set up CAS calls. CCS achieves greater call control because no contention exists between signaling and user traffic as it does with in-band CAS. Because the subscriber cannot generate particular signals intended for inter-switch (core network) signaling, CCS offers a greater degree of protection against fraud than analog CAS methods.

CCS has the following disadvantages in comparison to CAS:

- CCS links can be a single point of failure—a single link can control thousands of voice circuits, so if a link fails and no alternative routes are found, thousands of calls could be lost.
- There is no inherent testing of speech path by call set-up signaling, so elaborate Continuity Test procedures are required.

Chapter 3. The Role of SS7

The purpose of this chapter is to introduce Signaling System No. 7 (SS7/C7) and give the reader an indication of how it affects the lives of nearly two billion people globally. The chapter begins by providing a brief introduction to the major services that SS7/C7 provides and explains how the protocol has been and will continue to be a key enabler of new telecommunication services. It concludes with an explanation of why SS7/C7 is a cornerstone of convergence.

SS7/C7 is the protocol suite that is employed globally, across telecommunications networks, to provide signaling; it is also a private, "behind the scenes," packet-switched network, as well as a service platform. Being a signaling protocol, it provides the mechanisms to allow the telecommunication network elements to exchange control information.

AT&T developed SS7/C7 in 1975, and the *International Telegraph and Telephone Consultative Committee* (CCITT) [109] adopted it in 1980 as a worldwide standard. For more information on the standards bodies, see Chapter 2, "Standards." Over the past quarter of a century, SS7 has undergone a number of revisions and has been continually enhanced to support services that are taken for granted on a daily basis.

SS7/C7 is the key enabler of the public switched telephone network (PSTN), the integrated services digital network (ISDN), intelligent networks (INs), and public land mobile networks (PLMNs).

Each time you place and release a telephone call that extends beyond the local exchange, SS7/C7 signaling takes place to set up and reserve the dedicated network resources (trunk) for the call. At the end of the call, SS7/C7 takes action to return the resources to the network for future allocation.

TIP

Calls placed between subscribers who are connected to the same switch do not require the use of SS7/C7. These are known as intraoffice, intraexchange, or line-to-line calls.

Each time a cellular phone is powered up, SS7/C7-based transactions identify, authenticate, and register the subscriber. Before a cellular call can be made, further transactions check that the cellular phone is not stolen (network dependent option) and qualify permission to place the call (for example, the subscriber may be barred from International usage). In addition, the SS7/C7 network tracks the cellular subscriber to allow call delivery, as well as to allow a call that is already in progress to remain connected, even when the subscriber is mobile.

Although the average person typically uses SS7/C7 several times a day, it is largely unheard of by the general public because it is a "behind the scenes" private network—in stark contrast to IP. Another reason for its great transparency is its extreme reliability and resilience. For example, SS7/C7 equipment must make carrier grade quality standards—that is, 99.999 percent availability. The three prime ways it achieves an industry renowned robustness is by having a protocol that ensures reliable message delivery, self-healing capabilities, and an over-engineered physical network.

Typically, the links that comprise the network operate with a 20–40 percent loading and have full redundancy of network elements. SS7/C7 might well be the most robust and reliable network in existence.

SS7/C7 is possibly the most important element from a *quality of service* (QoS) perspective, as perceived by the subscriber.

NOTE

Here QoS refers to the quality of services as perceived by the subscriber. It should not be confused with QoS as it relates specifically to packet networks.

QoS is quickly becoming a key in differentiating between service providers. Customers are changing service providers at an increasing pace for QoS reasons, such as poor coverage, delays, dropped calls, incorrect billing, and other service-related impairments and faults. SS7/C7 impairments nearly always impact a subscriber's QoS directly. A complete loss of signaling means a complete network outage, be it a cellular or fixed-line network. Even a wrongly-provisioned screening rule at a SS7/C7 node in a cellular network can prohibit subscribers from roaming internationally or sending text messages. A loss of one signaling link could potentially bring down thousands of calls. For this reason, the SS7/C7 network has been designed to be extremely robust and resilient.

Impact of SS7 Network Failure

The critical nature of the SS7 network and the potential impact of failures was demonstrated in January 1990 when a failure in the SS7 software of an AT&T switching node rippled through over 100 switching nodes. The failure caused a nine-hour outage, affecting an estimated 60,000 people and costing in excess of 60 million dollars in lost revenue as estimated by AT&T.

Signaling System No. 7-Based Services

In addition to setting up and releasing calls, SS7/C7 is the workhorse behind a number of telecommunication services, including:

- Telephone-marketing numbers such as toll-free and freephone
- Televoting (mass calling)
- Single Directory Number
- *Enhanced 911* (E911)—used in the United States
- Supplementary services
- Custom local area signaling services (CLASS)
- Calling name (CNAM)
- Line information database (LIDB)
- Local number portability (LNP)
- Cellular network mobility management and roaming

  - Short Message Service (SMS)

  - *Enhanced Messaging Service* (EMS)— Ringtone, logo, and cellular game delivery

- Local exchange carrier (LEC) provisioned private virtual networks (PVNs)
- Do-not-call enforcement

The following sections describe these telecommunications services.

Telephone-Marketing Numbers

The most commonly used telephone-marketing numbers are *toll-free* calling numbers (800 calling), known as *freephone* (0800) in the United Kingdom. Because the call is free for the caller, these numbers can be used to win more business by increasing customer response. Telephone-marketing numbers also provide premium rate lines in which the subscriber is charged at a premium in exchange for desired content. Examples of such services include adult services and accurate road reports.

Another popular telephone-marketing number is *local call*, with which a call is charged as a local call even though the distance might be national. In recent years in the United Kingdom, marketing numbers that scarcely alter the call cost have been a popular means of masking geographical location. These numbers allow for a separation between the actual number and the advertised number.

Televoting

Televoting is a mass calling service that provides an easy method of surveying the public on any imaginable subject. The host (for example, a deejay at a radio station) presents specific questions and the caller uses a telephone keypad to select a choice; the caller's action adds to the vote for that particular choice. The conversation phase is usually limited to a simple, automated "thank you for…" phrase. Televoting can also be used in many other areas, such as responding to fundraising pleas and telephone-based competitions. A single night of televoting might result in 15 million calls [110]. Televoting services represent some of the most demanding—as well as lucrative—call scenarios in today's telephone networks. Revenue generation in this area is likely to grow as customers shift more toward an "interactive" experience, on par with convergence.

Single Directory Number

Another service that uses SS7/C7 and has been deployed in recent years is the single directory number, which allows a company with multiple offices or store locations to have a single directory number. After analyzing the calling party's number, the switch directs the call to a local branch or store.

Enhanced 911

E911, which is being deployed across some states in the United States, utilizes SS7 to transmit the number of the calling party, look up the corresponding address of the subscriber in a database, and transmit the information to the emergency dispatch operator to enable a faster response to emergencies. E911 might also provide other significant location information, such as the location of the nearest fire hydrant, and potentially the caller's key medical details. The *Federal Communications Commission* (FCC) also has a cellular 911 program in progress; in addition to providing the caller's telephone number, this program sends the geographical location of the antenna to which the caller is connected. Enhancement proposals are already underway to obtain more precise location information.

Supplementary Services

Supplementary services provide the subscribers with more than *plain old telephony service* (POTS), without requiring them to change their telephone handsets or access technology. Well-known supplementary services include three-way calling, *calling number display* (CND), call-waiting, and call forwarding. Note that the exact names of these services might differ, depending on the country and the operator.

Recently, supplementary services have been helpful in increasing operators' revenues since revenues against call minutes have been on the decline. Usually the subscriber must pay a fixed monthly or quarterly fee for a supplementary service.

Custom Local Area Signaling Services (CLASS)

Custom local area signaling services (CLASS) are an extension of supplementary services that employ the use of SS7 signaling between exchanges within a local geographical area. Information provided over SS7 links, such as the calling party number or the state of a subscriber line, enable more advanced services to be offered by service providers. A few examples of CLASS services include:

- **Call block**— Stops pre-specified calling party numbers from calling.
- **Distinctive ringing**— Provides a distinct ringing signal when an incoming call originates from a number on a predefined list. This feature is particularly beneficial to households with teenagers.
- **Priority ringing**— Provides a distinct ring when a call originates from a pre-specified numbers. If the called subscriber is busy and has *call waiting,* the subscriber receives a special tone indicating that a number on the priority list is calling.
- **Call completion to busy subscriber (CCBS)**— If a subscriber who has CCBS calls a party who is engaged in another call, the subscriber can activate CCBS with a single key or sequence. When activated, CCBS causes the calling party's phone to ring when the called party becomes available; when the calling party answers, the called party's phone automatically rings again. This feature saves the calling party from continuously attempting to place a call to a party is still unavailable.

Note that the exact names of these services might differ, depending on the country and the operator. In addition, the term "CLASS" is not used outside of North America.

Calling Name (CNAM)

Calling name (CNAM) is an increasingly popular database-driven service that is only available in the United States at this time. With this service, the called party receives the name of the person calling in addition to their number. The called party must have a compatible display box or telephone handset to use this service. The CNAM information is typically stored in regional telecommunications databases. SS7/C7 queries the database for the name based on the number and delivers the information to the called party's local switch.

Line Information Database (LIDB)

Line information database (LIDB) is a multipurpose database that stores valuable information about individual subscribers to provide feature-based services (it is only available in the United States at this time). Such information might include the subscriber's profile, name and address, and billing validation data. The name and address information can be used to power CNAM, for example. The billing validation data is used to support alternate billing services such as calling card, collect, and third number billing. Alternate billing services allow subscribers to bill calls to an account that is not necessarily associated with the originating line. For example, it can be used to validate a subscriber's calling card number that is stored in the LIDB, designating this as the means of payment. SS7/C7 is responsible for the real-time database query/response that is necessary to validate the calling card before progressing to the call setup phase.

Local Number Portability (LNP)

Local number portability (LNP) provides the option for subscribers to retain their telephone number when changing their telephone service. There are three phases of number portability:

- Service Provider Portability
- Service Portability
- Location Portability

The various phases of LNP are discussed in more detail in Chapter 11, "Intelligent Networks."

The FCC mandated this feature for fixed-line carriers in the United States as part of the Telecommunications Act of 1996; later that same year, the act was also clarified to cover cellular carriers.

LNP is primarily aimed at stimulating competition among providers by removing the personal inconvenience of changing phone numbers when changing service providers. For example, many businesses and individuals spend relatively large sums of money to print their phone numbers on business cards, letterheads, and other correspondence items. Without LNP, people would have to reprint and redistribute these materials more often. This contributes to the inconvenience and detracts from the profitability of changing the telephone number, thereby making changing providers far more prohibitive.

Since telephone networks route calls based on service provider and geographic numbering plan information, SS7/C7 must figure out where the ported number's new terminating switch is by performing additional signaling before setting the call up. This step should add only a second to the call overhead setup; however, it is a technically challenging network change because it complicates the process by which SS7/C7 establishes a call behind the scenes. This process is further discussed in Chapter 8, "ISDN User Part (ISUP)."

2nd and 3rd Generation Cellular Networks

Cellular networks use SS7/C7 for the same reasons they use fixed line networks, but they place much higher signaling demands on the network because of subscriber mobility. All

cellular networks, from 2G (GSM, ANSI-41, and even PDC, which is used in Japan) to 3G (UMTS and cdma2000), use SS7/C7 for call delivery, supplementary services, roaming, mobility management, prepaid, and subscriber authentication. For more information, see Chapter 13, "GSM and ANSI-41 Mobile Application Part (MAP)."

Short Message Service (SMS)

Short Message Service (SMS) forms part of the GSM specifications and allows two-way transmission of alphanumeric text between GSM subscribers. Although it is just now catching on in North America, SMS has been an unexpected and huge revenue source for operators around the world. Originally, SMS messages could be no longer than 160 alphanumeric characters. Many handsets now offer concatenated SMS, which allows users to send and receive messages up to 459 characters (this uses EMS described below). Cellular operators usually use SMS to alert the subscribers that they have voice mail, or to educate them on how to use network services when they have roamed onto another network. Third party companies offer the additional delivery services of sending SMS-to-fax, fax-to-SMS, SMS-to-e-mail, e-mail-to-SMS, SMS-to-web, web-to-SMS, and SMS notifications of the arrival of new e-mail.

Some European (Spain, Ireland, and Germany, for example) and Asian countries (the Philippines, for example) are rolling out fixed-line SMS, which allows users to send SMS through their fixed phone line to cell phones and vice versa, as well as to other fixed-line SMS-enabled phones, fax machines, e-mail, and specialized web pages. Thus far, each European rollout has also offered SMS-to-voice mail. If a caller sends a text message to a subscriber without fixed-line SMS facility, the SMS is speech-synthesized to the subscriber's and their voice mailbox. Fixed-line SMS requires compatible phones, which are becoming readily available.

SMS is carried on the SS7/C7 network, and it makes use of SS7/C7 for the required signaling procedures. For more information, see Chapter 13, "GSM and ANSI-41 Mobile Application Part (MAP)."

Enhanced Messaging Service (EMS)

Enhanced Messaging Service (*EMS*) adds new functionality to the SMS service in the form of pictures, animations, sound, and formatted text. EMS uses existing SMS infrastructure and consists largely of header changes made to a standard SMS message. Since EMS is simply an enhanced SMS service, it uses the SS7/C7 network in the same way; the SS7/C7 network carries it, and it uses SS7/C7 for the required signaling procedures.

EMS allows users to obtain new ring tones, screensavers, pictures, and animations for their cell phones either by swapping with friends or purchasing them online.

Operators have recently begun using EMS for downloading games (from classics like Asteroids, to newer games like Prince of Persia), which can be purchased from operator web sites.

Private Virtual Networks

Although the *private virtual networks* concept is not new, SS7/C7 makes it possible for a *Local exchange carrier* (LEC) to offer the service. The customer receives PVNs, which are exactly like leased (private) lines except that the network does not allocate dedicated physical resources. Instead, SS7/C7 signaling (and a connected database) monitors the "private customer" line. The customer has all the features of a leased-line service as well as additional features, such as the ability to request extra services ad hoc and to tailor the service to choose the cheapest *inter-exchange carrier* (IC), depending on the time of day, day or week, or distance between the two parties.

Do-Not-Call Enforcement

In the United States, federal and state laws have already mandated do-not-call lists [108] in over half the states, and all states are expected to follow suit. These laws restrict organizations (typically telemarketers) from cold-calling individuals. To comply with these laws, SS7 can be used to query state and federal do-not-call lists (which are stored on a database) each time a telemarketer makes an outbound call. If the number is on a do-not-call list, the call is automatically blocked and an appropriate announcement is played to the marketer.

Signaling System No. 7: The Key to Convergence

Telecommunications network operators can realize increased investment returns by marrying existing SS7/C7 and intelligent networking infrastructures with Internet and other data-centric technologies. SS7/C7 is a key protocol for bridging the telecom and datacom worlds. telecom & datacom

The following sections describe the exemplar hybrid network services that SS7/C7 enable:

- Internet Call Waiting
- Internet Calling Name Services
- Click-to-Dial Applications
- Web-Browser-Based Telecommunication Services
- WLAN "Hotspot" Billing
- Location-Based Games

Internet Call Waiting and Internet Calling Name Services

Internet call waiting is a software solution that alerts online Internet users with a call-waiting message on their computer screens when a telephone call enters the same phone line they use for their Internet service. The user can then send the call to voice mail, accept the call, or reject it.

Some providers linking it to CNAM, as mentioned in Calling Name (CNAM), have enhanced the Internet call-waiting service. This service is known as Internet calling name service, and it provides the calling party's name and number.

Click-to-Dial Applications

Click-to-dial applications are another SS7-IP growth area. An example of a click-to-dial application is the ability to click a person's telephone number in an email signature to place a call. These types of services are particularly beneficial to subscribers because they do not require them to change their equipment or access technologies; a POTS and a traditional handset are the only requirements.

Web-Browser-Based of Telecommunication Services

Over the coming decade, we are likely to witness an increase in web based telecommunications services. An example is customer self-provisioning via the Internet, a practice that has been in the marketplace for some time and is likely to increase in both complexity and usage. A customer can already assign himself a premium or toll-free "number for life" via the Internet. The customer can subsequently use a Web interface to change the destination number it points to at will, so that during the day it points to the customer's office phone, and in the evening it points to the customer's cell phone, and so forth.

Another example is the "call me" service, which allows a customer to navigate a Web page to arrange a callback from a department, rather than navigating interactive voice response (IVR) systems through the use of voice prompts and a touch-tone phone.

The potential extends far beyond traditional telecommunications services, to the point where the distinction between Web and telecommunications services is blurred. An example of such an enabling technology is Voice Extensible Markup Language (VoiceXML), which extends Web applications to telephones and shields application authors from low-level, platform-specific *interactive voice response* (IVR) and call control details.

The marriage is not only between SS7/C7, the Internet, and fixed-line networks—it also extends to cellular networks. Plans are underway to put the location-based information and signaling found in cellular networks into hybrid use. For example, Web-based messenger services could access cellular network *home location registers* (HLRs) to enable a user to locate a friend or relative in terms of real-time geographic location.

WLAN "Hotspot" Billing

SS7/C7 has recently begun playing a role in the marriage of wireless (WLANs) and cellular networks. A subscriber can use a cellular subscriber identity module (SIM) card for authentication and billing purposes from a WLAN hotspot. For example, if a subscriber is at a café with WLAN facilities (typically wi-fi), the subscriber can request permission to use the service via a laptop screen. This request triggers a short cellular call to authenticate the subscriber (using SS7/C7 signaling). The usage is then conveniently billed to the subscriber's cellular phone bill.

NOTE

A SIM is used in 2nd generation cellular networks based on GSM, and on 2.5/3G networks as defined by 3GPP. A SIM contains the subscriber's identity so that the subscriber can change cellular equipment freely by simply changing the SIM card over to the new device. This

means that the subscriber can plug the SIM into a new cellular handset and the number "transfers" to that handset, along with the billing.

Location-Based Games

SS7/C7 is not only used to deliver games to cell phones, but it also plays a role in the creation of a new genre of location-based games and entertainment. Cellular games incorporate the player's location using SS7/C7 to provide mobility information a dedicated web site as a central point. Some of the games that are emerging at the time of this writing are using global positioning system (GPS), WLAN support, and built-in instant messaging capabilities (to help tease your opponents) to blend higher location accuracy.

Summary

This chapter has shown that, although it is transparent, SS7/C7 plays a role in the lives of virtually every individual in developed countries. It is also the key to new, revenue-generating services and is crucial to the QoS as perceived by subscribers—both of which lie at the very heart of success in a fiercely competitive telecommunications market. Furthermore SS7/C7 is a common thread that ties fixed-line, cellular, and IP networks together, and it is a key enabler for the convergence of the telecommunications and data communications industries.

Chapter 4. SS7 Network Architecture and Protocols Introduction

The *International Telecommunication Union (ITU)* is the international governing body for Signaling System No. 7. More specifically, it is governed by the Telecommunication Standardization Sector of the ITU (ITU-TS or ITU-T for short). Formerly it was governed by the ITU's Consultative Committee for International Telegraph and Telephone (CCITT) subcommittee until that was disbanded in 1992 as part of a process to speed up the production of recommendations (as well as other organization changes). See Chapter 2, "Standards," for more information on standards-making bodies.

Signaling System No. 7 is more commonly known by the acronyms SS7 and C7. Strictly speaking, the term C7 (or, less commonly, CCS7) refers to the international Signaling System No. 7 network protocols specified by the ITU-T recommendations as well as national or regional variants defined within the framework provided by the ITU-T. The term C7 originates from the former title found on the specifications—CCITT Signaling System No. 7. The term SS7 tends to specifically refer to the North American regional standards produced by Telcordia (formerly known as Bell Communications Research or Bellcore) and the American National Standards Institute (ANSI). The North American standards themselves are based on the ITU-T recommendations but have been tailored outside the provided framework. The differences between ITU and Telcordia/ANSI are largely subtle at the lower layers. Interaction between ANSI and ITU-T networks is made challenging by different implementations of higher-layer protocols and procedures.

For the purpose of this book, we will use the term SS7 to refer generically to any Signaling System No. 7 protocol, regardless of its origin or demographics. An overview of SS7 by the ITU-T can be found in recommendation Q.700 [111], and a similar overview of SS7 by ANSI can be found in T1.110 [112].

Chapter 3, "The Role of SS7," provides a comprehensive list of the functions and services afforded by SS7. These can be summarized as follows:

- Setting up and tearing down circuit-switched connections, such as telephone calls made over both cellular and fixed-line.
- Advanced network features such as those offered by supplementary services (calling name/number presentation, Automatic Callback, and so on).
- Mobility management in cellular networks, which permits subscribers to move geographically while remaining attached to the network, even while an active call is in place. This is the central function of a cellular network.
- *Short Message Service (SMS)* and *Enhanced Messaging Service (EMS)*, where SS7 is used not only for signaling but also for content transport of alphanumeric text.
- Support for *Intelligent Network (IN)* services such as toll-free (800) calling.
- Support for ISDN.
- *Local Number Portability (LNP)* to allow subscribers to change their service, service provider, and location without needing to change their telephone number.

After reading the preceding chapters, you know that signaling serves the requirements of the telecommunications service being delivered; it is not an end in itself. Signaling enables services within the network.

This chapter makes you familiar with the SS7 network, protocols, fundamental concepts, and terminology so that the topics covered in the rest of the book will be more accessible if you're unfamiliar with the subject. This chapter begins with a brief description of pre-SS7 systems and SS7 history. The chapter then presents the protocol stack, showing how SS7 protocols fit together. It concludes with a discussion of the relevant protocols.

Pre-SS7 Systems

The following are the main systems that preceded SS7:

- CCITT R1 (regional 1) was deployed only on a national level. R1 is a *Channel Associated Signaling (CAS)* system that was employed in the U.S. and Japan. It uses multifrequency (MF) tones for signaling. It is no longer in general operation, although some remnants might remain in the network.
- CCITT R2 (regional 2) was deployed only on a national level. R2 is a CAS system that was employed in Europe and most other countries. It used *Multifrequency Compelled (MFC)* for signaling; it compelled the receiver to acknowledge a pair of tones before sending the next pair. It is no longer in general operation, although some remnants might remain in the network.
- Signaling systems that have been deployed for both national and international (between international switches) signaling have progressed from CCITT #5 (C5) to CCITT #6 (C6) and finally to CCITT #7 (C7):

  - C5 (CCITT Signaling System No. 5) is a CAS system standardized in 1964 that has found widespread use in international signaling. It is still in use today on a number of

international interfaces. National implementations are now scarce, except in less-developed regions of the world, such as Africa, which makes extensive use of the protocol. C5 can be used in both analog and digital environments. In an analog setting, it uses tones for signaling. In a digital setting, a digital representation of the tone is sent instead (a pulse code modulation [PCM] sample).

- C6 (CCITT Signaling System No. 6), also called SS6, was the first system to employ *Common Channel Signaling* (CCS). It was standardized in 1972. (CAS and CCS are explained in Chapter 1, "The Evolution of Signaling.") C6 was a pre-OSI model and as such had a monolithic structure as opposed to a layered one. C6 was a precursor to C7 and included the use of data links to carry signaling in the form of packets. It had error correction/detection mechanisms. It employed a common signaling channel to control a large number of speech circuits, and it had self-governing network management procedures. C6 had a number of advantages over C5, including improvements in post-dial delay and the ability to reject calls with a cause code. The use of locally mapped cause codes allowed international callers to hear announcements in their own language. Although C6 was designed for the international network, it was not as widely deployed as C5. However, it was nationalized for the U.S. network and was deployed quite extensively under the name *Common Channel Interoffice Signaling System* 6 *(CCIS6)* in the AT&T network. C6 was introduced into the Bell system in the U.S. in 1976, and soon after, Canada. All deployments have now been replaced by SS7.

The next section provides a brief history of SS7.

History of SS7

The first specification (called a *recommendation* by the CCITT/ITU-T) of CCITT Signaling System No. 7 was published in 1980 in the form of the CCITT *yellow book* recommendations. After the yellow book recommendations, CCITT recommendations were approved at the end of a four-year study period. They were published in a colored book representing that study period.

Table 4-1 provides an evolutionary time line of CCITT/ITU-T SS7.

Table 4-1. CCITT/ITU-T SS7 Timeline

Year	Publication	Protocols Revised or Added
1980	CCITT Yellow Book	MTP2, MTP3, and TUP, first publication.
1984	CCITT Red Book	MTP2, MTP3, and TUP revised. SCCP and ISUP added.
1988	CCITT Blue Book	MTP2, MTP3, TUP, and ISUP revised. ISUP supplementary services and TCAP added.
1992	ITU-T Q.767	International ISUP, first publication.
1993	ITU-T "White Book 93"	ISUP revised.
1996	ITU-T "White Book 96"	MTP3 revised.

Table 4-1. CCITT/ITU-T SS7 Timeline

Year	Publication	Protocols Revised or Added
1997	ITU-T "White Book 97"	ISUP revised.
1999	ITU-T "White Book 99"	ISUP revised.

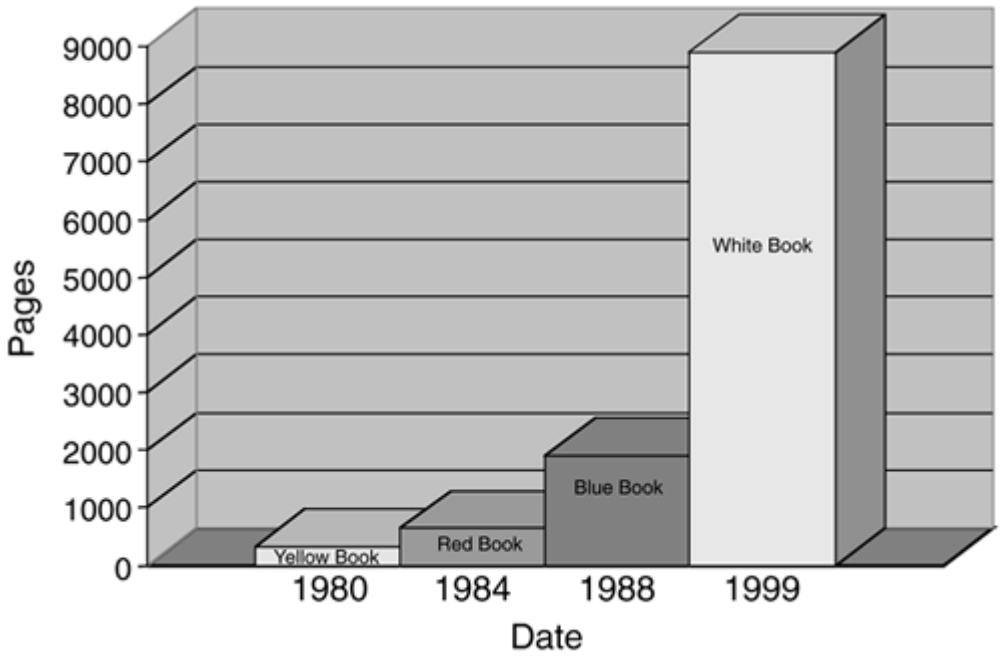
Under the CCITT publishing mechanism, the color referred to a published set of recommendations—that is, all protocols were published at the same time. The printed matter had the appropriate colored cover, and the published title contained the color name. When the ITU-T took over from the CCITT, it produced single booklets for each protocol instead of producing en bloc publications as had been the case under the supervision of the CCITT. Under the new mechanism, the color scheme was dropped. As a result, the ITU-T publications came to be known as "White Book" editions, because no color was specified, and the resulting publications had white covers. Because these publications do not refer to a color, you have to qualify the term "White Book" with the year of publication.

As Table 4-1 shows, when SS7 was first published, the protocol stack consisted of only the Message Transfer Part 2 (MTP2), Message Transfer Part 3 (MTP3), and Telephony User Part (TUP) protocols. On first publication, these were still somewhat immature. It was not until the later Red and Blue book editions that the protocol was considered mature. Since then, the SS7 protocols have been enhanced, and new protocols have been added as required.

Figure 4-1 shows how many pages the ITU-T SS7 specifications contained in each year. In 1980, there were a total of 320 pages, in 1984 a total of 641 pages, in 1988 a total of 1900 pages, and in 1999 approximately 9000 pages.

Figure 4-1. How Many Pages the ITU C7 Specifications Covered Based on Year (Source: ITU [Modified])

[View full size image]



The following section introduces the SS7 network architecture.

SS7 Network Architecture

SS7 can employ different types of signaling network structures. The choice between these different structures can be influenced by factors such as administrative aspects and the structure of the telecommunication network to be served by the signaling system.

The worldwide signaling network has two functionally independent levels:

- International
- National

This structure makes possible a clear division of responsibility for signaling network management. It also lets numbering plans of SS7 nodes belonging to the international network and the different national networks be independent of one another.

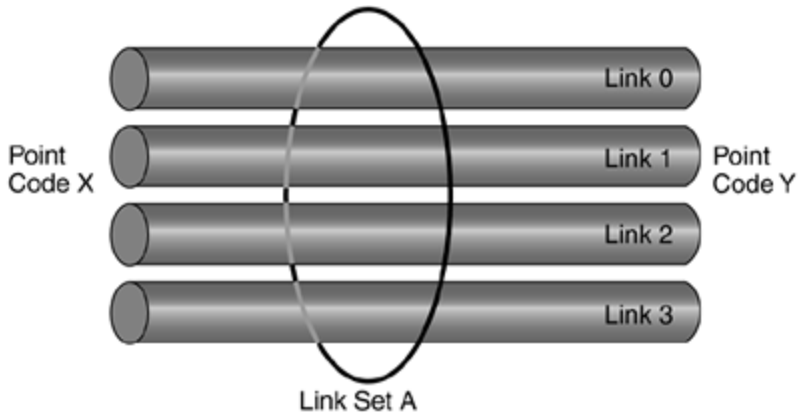
SS7 network nodes are called signaling points (SPs). Each SP is addressed by an integer called a point code (PC). The international network uses a 14-bit PC. The national networks also use a 14-bit PC—except North America and China, which use an incompatible 24-bit PC, and Japan, which uses a 16-bit PC. The national PC is unique only within a particular operator's national network. International PCs are unique only within the international network. Other operator networks (if they exist) within a country also could have the same PC and also might share the same PC as that used on the international network. Therefore, additional routing information is provided so that the PC can be interpreted correctly—that is, as an international network, as its own national network, or as another operator's national network. The structure of point codes is described in Chapter 7, "Message Transfer Part 3 (MTP3)."

Signaling Links and Linksets

SPs are connected to each other by signaling links over which signaling takes place. The bandwidth of a signaling link is normally 64 kilobits per second (kbps). Because of legacy reasons, however, some links in North America might have an effective rate of 56 kbps. In recent years, high-speed links have been introduced that use an entire 1.544 Mbps T1 carrier for signaling. Links are typically engineered to carry only 25 to 40 percent of their capacity so that in case of a failure, one link can carry the load of two.

To provide more bandwidth and/or for redundancy, up to 16 links between two SPs can be used. Links between two SPs are logically grouped for administrative and load-sharing reasons. A logical group of links between two SP is called a *linkset*. Figure 4-2shows four links in a linkset.

Figure 4-2. Four Links in a Linkset Between SPs



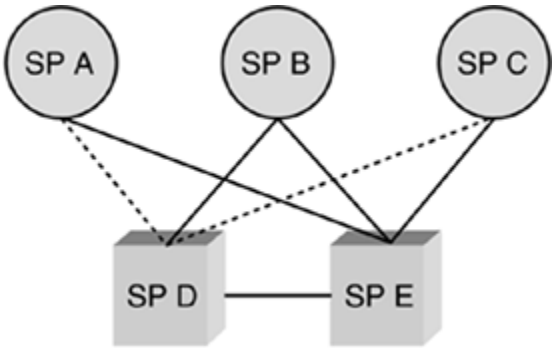
A number of linksets that may be used to reach a particular destination can be grouped logically to form a *combined linkset*. For each combined linkset that an individual linkset is a member of, it may be assigned different priority levels relative to other linksets in each combined linkset.

A group of links within a linkset that have the same characteristics (data rate, terrestrial/satellite, and so on) are called a *link group*. Normally the links in a linkset have the same characteristics, so the term *link group* can be synonymous with *linkset*.

Routes and Routesets

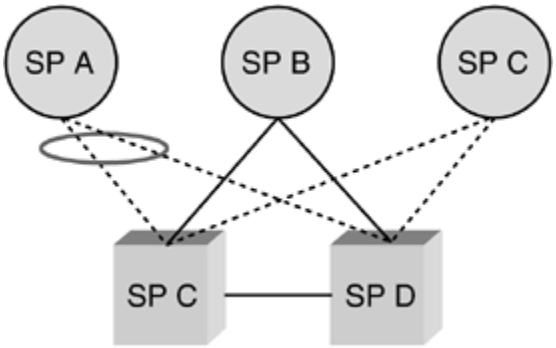
SS7 routes are statically provisioned at each SP. There are no mechanisms for route discovery. A *route* is defined as a preprovisioned path between source and destination for a particular relation. Figure 4-3 shows a route from SP A to SP C.

Figure 4-3. Route from SP A to SP C



All the preprovisioned routes to a particular SP destination are called the *routeset*. Figure 4-4 shows a routeset for SSP C consisting of two routes.

Figure 4-4. Routeset from SP A to SP C



The following section discusses the SP types.

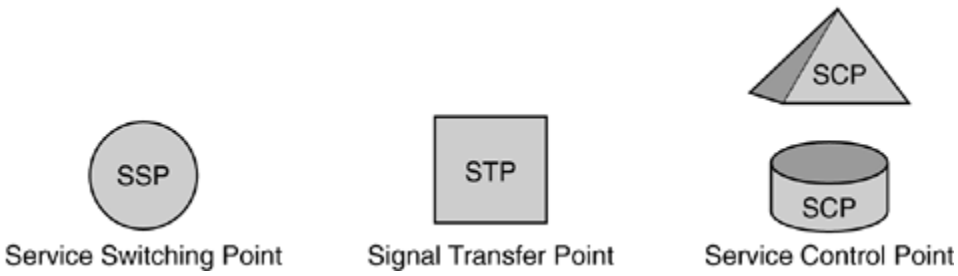
Node Types

There are three different types of SP (that is, SS7 node):

- Signal Transfer Point
- Service Switching Point
- Service Control Point

Figure 4-5 graphically represents these nodes.

Figure 4-5. SS7 Node Types



The SPs differ in the functions that they perform, as described in the following sections.

Signal Transfer Point

A Signal Transfer Point (STP) is responsible for the transfer of SS7 messages between other SS7 nodes, acting somewhat like a router in an IP network.

They are intermediate SPs

An STP is neither the ultimate source nor the destination for most signaling messages. Generally, messages are received on one signaling link and are transferred out another. The only messages that are not simply transferred are related to network management and global title translation. These two functions are discussed more in Chapters 7 and 9. STPs route each incoming message to an outgoing signaling link based on routing information contained in the SS7 message. Specifically, this is the information found in the MTP3 routing label, as described in Chapter 7.

Additionally, standalone STPs often can screen SS7 messages, acting as a firewall. Such usage is described in Chapter 15, "SS7/C7 Security and Monitoring."

An STP can exist in one of two forms:

- Standalone STP
- Integrated STP (SP with STP)

Standalone STPs are normally deployed in "mated" pairs for the purposes of redundancy. Under normal operation, the mated pair shares the load. If one of the STPs fails or isolation occurs because of signaling link failure, the other STP takes the full load until the problem with its mate has been rectified.

Integrated STPs combine the functionality of an SSP and an STP. They are both the source and destination for MTP user traffic. They also can transfer incoming messages to other nodes.

Service Switching Point

A Service Switching Point (SSP) is a voice switch that incorporates SS7 functionality. It processes voice-band traffic (voice, fax, modem, and so forth) and performs SS7 signaling. All switches with SS7 functionality are considered SSPs regardless of whether they are local switches (known in North America as an end office) or tandem switches.

An SSP can originate and terminate messages, but it cannot transfer them. If a message is received with a point code that does not match the point code of the receiving SSP, the message is discarded.

Service Control Point

A Service Control Point (SCP) acts as an interface between telecommunications databases and the SS7 network. Telephone companies and other telecommunication service providers employ a number of databases that can be queried for service data for the provision of services. Typically the request (commonly called a query) originates at an SSP. A popular example is freephone calling (known as toll-free in North America). The SCP provides the routing number (translates the toll-free number to a routable number) to the SSP to allow the call to be completed. For more information, see Chapter 11, "Intelligent Networks (IN)."

SCPs form the means to provide the core functionality of cellular networks, which is subscriber mobility. Certain cellular databases (called registers) are used to keep track of the subscriber's location so that incoming calls may be delivered. Other telecommunication databases include those used for calling card validation (access card, credit card), calling name display (CNAM), and LNP.

SCPs used for large revenue-generating services are usually deployed in pairs and are geographically separated for redundancy. Unless there is a failure, the load is typically shared between two *mated* SCPs. If failure occurs in one of the SCPs, the other one should be able to take the load of both until normal operation resumes.

Queries/responses are normally routed through the mated pair of STPs that services that particular SCP, particularly in North America.

See Chapters 10, "Transaction Capabilities Application Part (TCAP)," and 11, "Intelligent Networks (IN)," for more information on the use of SCPs within both fixed-line and cellular networks. See Chapters 12, "Cellular Networks," and 13, "GSM and ANSI-41 Mobile Application Part (MAP)," for specific information on the use of SCPs within cellular networks.

The following section introduces the concept of link types.

Link Types

Signaling links can be referenced differently depending on where they are in the network. Although different references can be used, you should understand that the link's physical characteristics remain the same. The references to link types A through E are applicable only where standalone STPs are present, so the references are more applicable to the North American market.

Six different link references exist:

- Access links (A links)
- Crossover links (C links)
- Bridge links (B links)
- Diagonal links (D links)
- Extended links (E links)
- Fully associated links (F links)

The following sections cover each link reference in more detail.

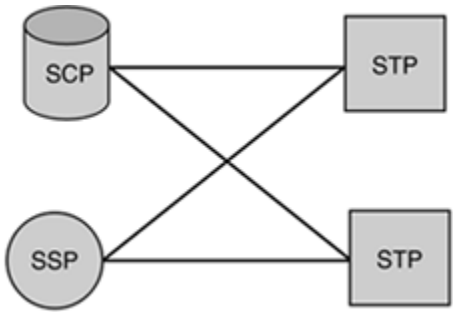
NOTE

In the figures in the sections covering the different link references, dotted lines represent the actual link being discussed, and solid lines add network infrastructure to provide necessary context for the discussion.

Access Links (A Links)

Access links (A links), shown in Figure 4-6, provide access to the network. They connect "outer" SPs (SSPs or SCPs) to the STP backbone. A links connect SSPs and SCPs to their serving STP or STP mated pair.

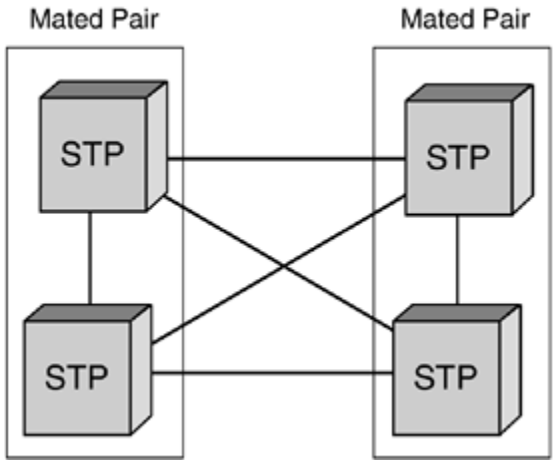
Figure 4-6. A Links



Cross Links (C Links)

Cross links (C links), shown in Figure 4-7, are used to connect two STPs to form a mated pair—that is, a pair linked such that if one fails, the other takes the load of both.

Figure 4-7. C Links



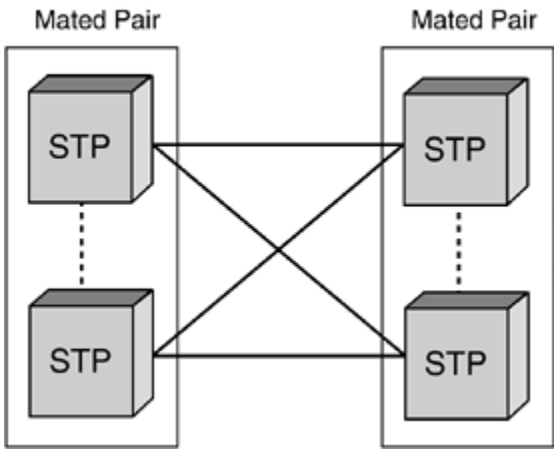
C links are used to carry MTP user traffic only when no other route is available to reach an intended destination. Under normal conditions, they are used only to carry network management messages.

Bridge Links (B Links)

Bridge links (B links) are used to connect mated pairs of STPs to each other across different regions within a network at the same hierarchical level. These links help form the backbone of the SS7 network. B links are normally deployed in link quad configuration between mated pairs for redundancy.

Figure 4-8 shows two sets of mated pairs of B links.

Figure 4-8. B Links

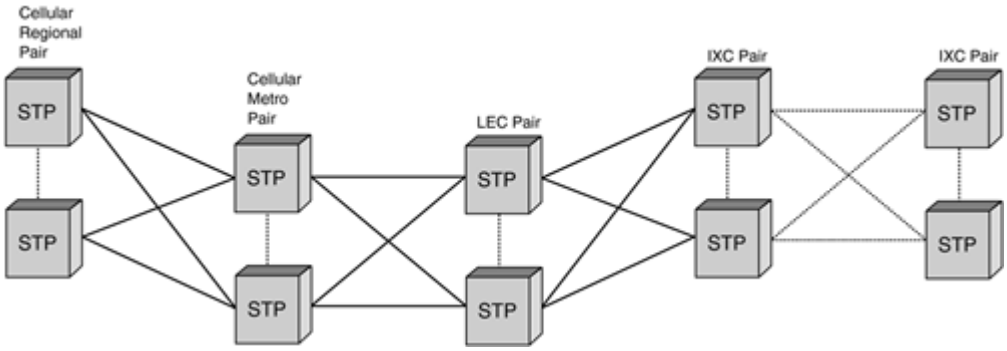


Diagonal Links (D Links)

Diagonal links (D links), shown in Figure 4-9, are the same as B links in that they connect mated STP pairs.

Figure 4-9. D Links

[View full size image]

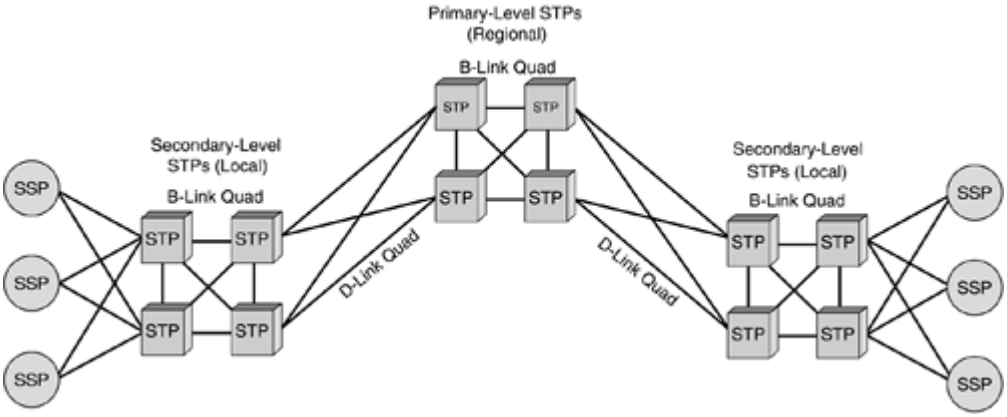


The difference is that they connect mated STP pairs that belong to different hierarchical levels or to different networks altogether. For example, they may connect an interexchange carrier (IXC) STP pair to a local exchange carrier (LEC) STP pair or a cellular regional STP pair to a cellular metro STP pair.

As mentioned, B and D links differ in that D links refer specifically to links that are used either between different networks and/or hierarchical levels, as shown in Figure 4-10.

Figure 4-10. Existence of an STP Backbone and STP Hierarchy

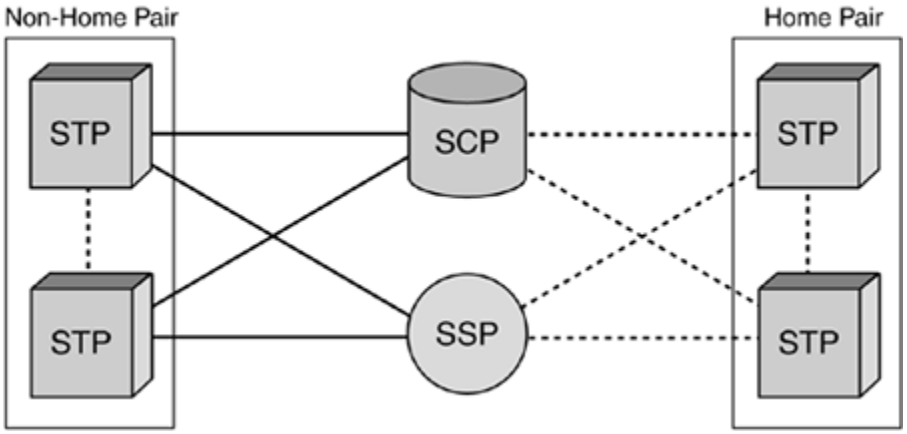
[View full size image]



Extended Links (E Links)

Extended links (E links), shown in Figure 4-11, connect SSPs and SCPs to an STP pair, as with A links, except that the pair they connect to is not the normal home pair. Instead, E links connect to a nonhome STP pair. They are also called alternate access (AA) links. E links are used to provide additional reliability or, in some cases, to offload signaling traffic from the home STP pair in high-traffic corridors. For example, an SSP serving national government agencies or emergency services might use E links to provide additional alternate routing because of the criticality of service.

Figure 4-11. E Links



Fully-Associated Links (F Links)

Fully-associated links (F links), shown in Figure 4-12, are used to connect network SSPs and/or SCPs directly to each other without using STPs. The most common application of this type of link is in metropolitan areas. F links can establish direct connectivity between all switches in the area for trunk signaling and Custom Local Area Signaling Service (CLASS), or to their corresponding SCPs.

Figure 4-12. F Links

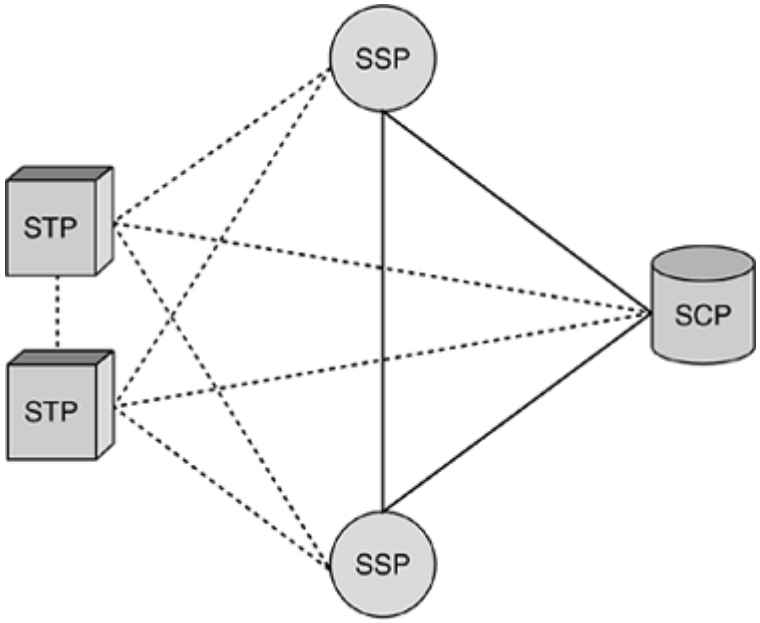
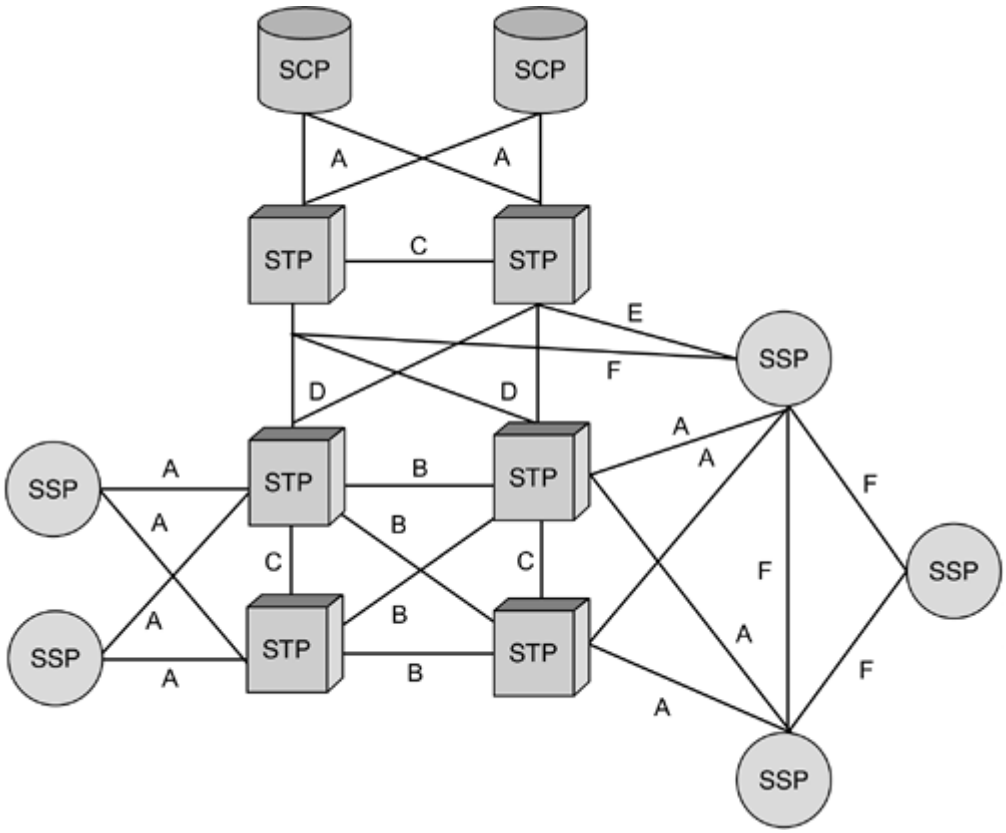


Figure 4-13 shows an SS7 network segment. In reality, there would be several factors more SSPs than STPs.

Figure 4-13. SS7 Network Segment



Signaling Modes

The signaling relationship that exists between two communicating SS7 nodes is called the signaling mode. The two modes of signaling are associated signaling and quasi-associated signaling. When the destination of an SS7 message is directly connected by a linkset, the *associated* signaling mode is being used. In other words, the source and destination nodes are directly connected by a single linkset. When the message must pass over two or more linksets and through an intermediate node, the *quasi-associated*mode of signaling is being used.

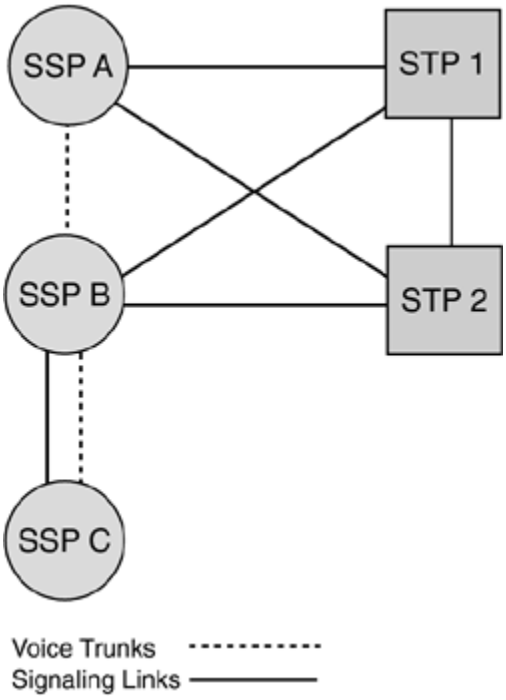
It's easier to understand the signaling mode if you examine the relationship of the point codes between the source and destination node. When using the associated mode of signaling, the Destination Point Code (DPC) of a message being sent matches the PC of the node at the far end of the linkset, usually referred to as the far-end PC or adjacent PC. When quasi-associated signaling is used, the DPC does not match the PC at the far end of the connected linkset. Quasi-associated signaling requires the use of an STP as the intermediate node because an SSP cannot transfer messages.

In Figure 4-14, the signaling relationships between each of the nodes are as follows:

- SSP A to SSP B uses quasi-associated signaling.

- SSP B to SSP C uses associated signaling.
- STP 1 and STP 2 use associated signaling to SSP A, SSP B, and each other.

Figure 4-14. SS7 Signaling Modes



As you can see from Figure 4-14, associated signaling is used between nodes that are directly connected by a single linkset, and quasi-associated signaling is used when an intermediate node is used. Notice that SSP C is only connected to SSP B using an F link. It is not connected to any other SS7 nodes in the figure.

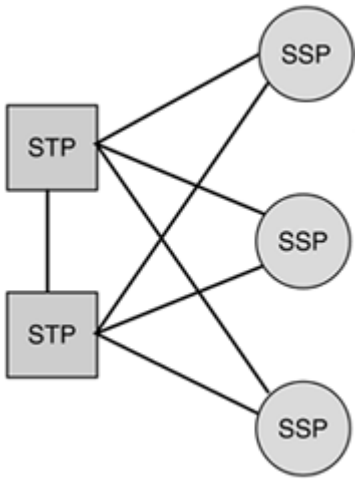
When discussing the signaling mode in relation to the voice trunks shown between the SSPs, the signaling and voice trunks follow the same path when associated signaling is used. They take separate paths when quasi-associated signaling is used. You can see from Figure 4-14 that the signaling between SSP B and SSP C follows the same path (associated mode) as the voice trunks, while the signaling between SSP A and SSP B does not follow the same path as the voice trunks.

Signaling Network Structure

Standalone STPs are prevalent in North America because they are used in this region to form the backbone of the SS7 network. Attached to this backbone are the SSPs and SCPs. Each SSP and SCP is assigned a "home pair" of STPs that it is directly connected to. The network of STPs can be considered an overlay onto the telecommunications network—a packet-switched data communications network that acts as the nervous system of the

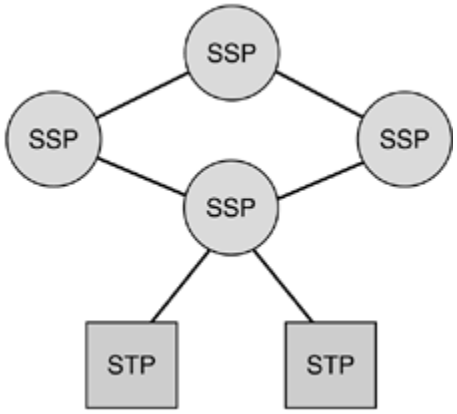
telecommunications network. Figure 4-15 shows a typical example of how SSPs are interconnected with the STP network in North America.

Figure 4-15. Typical Example of North American SSP Interconnections



STPs are not as common outside North America. Standalone STPs typically are used only between network operators and/or for applications involving the transfer of noncircuit-related signaling. In these regions, most SSPs have direct signaling link connections to other SSPs to which they have direct trunk connections. Figure 4-16 shows an example of this type of network with most SSPs directly connected by signaling links.

Figure 4-16. Typical Example of SSP Interconnections in Most Areas Outside North America

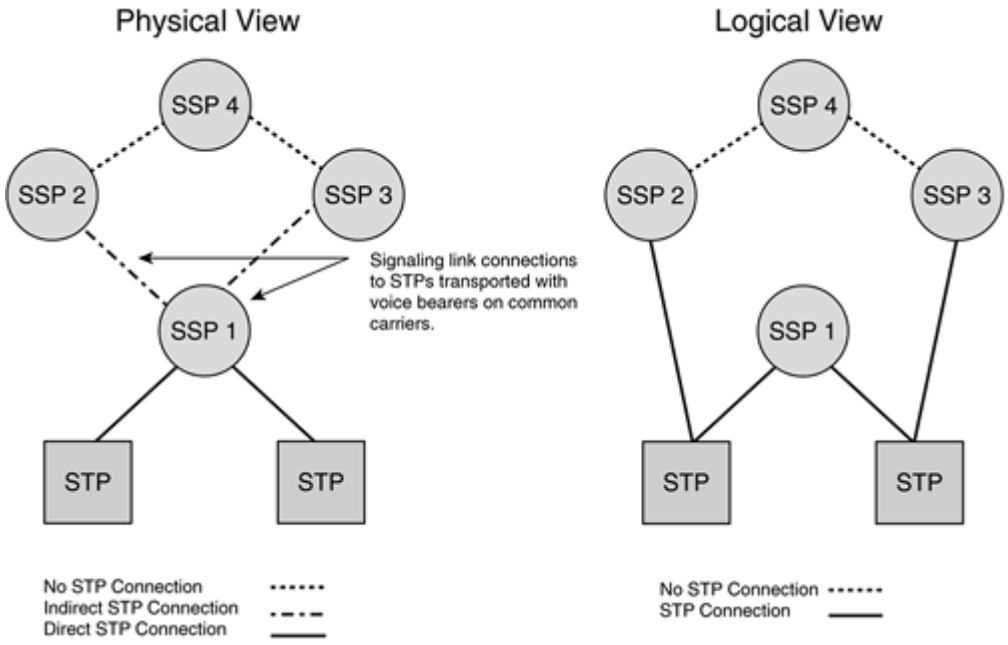


SSPs often have indirect physical connections to STPs, made through other SSPs in the network. These are usually implemented as nailed-up connections, such as through a Digital Access Cross-Connect System or other means of establishing a semipermanent connection. Logically, these SSPs are directly connected to the STP. The signaling link occupies a digital time slot on the same physical medium as the circuit-switched traffic. The SSPs that provide physical interconnection between other SSPs and an STP do not "transfer" messages as an STP function. They only provide physical connectivity of the signaling links between T1/E1

carriers to reach the STP. Figure 4-17 shows an example of a network with no STP connection, direct connections, and nondirect connections. SSP 1 is directly connected to an STP pair. SSP 4 uses direct signaling links to SSP 2 and SSP 3, where it also has direct trunks. It has no STP connection at all. SSP 2 and SSP 3 are connected to the STP pair via nailed-up connections at SSP 1.

Figure 4-17. Example of Direct and Indirect SSP Interconnections to STPs

[View full size image]



Normally within networks that do not use STPs, circuit-related (call-related) signaling takes the same path through the network as user traffic because there is no physical need to take a different route. This mode of operation is called *associated signaling* and is prevalent outside North America. Referring back to Figure 4-14, both the user traffic and the signaling take the same path between SSP B and SSP C.

Because standalone STPs are used to form the SS7 backbone within North America, and standalone STPs do not support user traffic switching, the SSP's signaling mode is usually quasi-associated, as illustrated between SSP A and SSP B in Figure 4-14.

In certain circumstances, the SSP uses associated signaling within North America. A great deal of signaling traffic might exist between two SSPs, so it might make more sense to place a signaling link directly between them rather than to force all signaling through an STP.

SS7 Protocol Overview

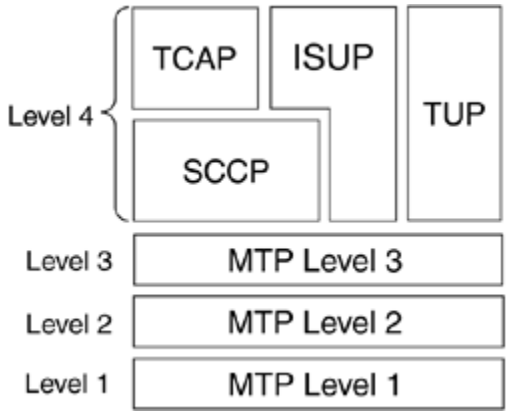
The number of possible protocol stack combinations is growing. It depends on whether SS7 is used for cellular-specific services or intelligent network services, whether transportation is over IP or is controlling broadband ATM networks instead of time-division multiplexing

(TDM) networks, and so forth. This requires coining a new term—traditional SS7—to refer to a stack consisting of the protocols widely deployed from the 1980s to the present:

- Message Transfer Parts (MTP 1, 2, and 3)
- Signaling Connection Control Part (SCCP)
- Transaction Capabilities Application Part (TCAP)
- Telephony User Part (TUP)
- ISDN User Part (ISUP)

Figure 4-18 shows a common introductory SS7 stack.

Figure 4-18. Introductory SS7 Protocol Stack



Such a stack uses TDM for transport. This book focuses on traditional SS7 because that is what is implemented. Newer implementations are beginning to appear that use different transport means such as IP and that have associated new protocols to deal with the revised transport.

The SS7 physical layer is called MTP level 1 (MTP1), the data link layer is called MTP level 2 (MTP2), and the network layer is called MTP level 3 (MTP3). Collectively they are called the *Message Transfer Part (MTP)*. The MTP protocol is SS7's native means of packet transport. In recent years there has been an interest in the facility to transport SS7 signaling over IP instead of using SS7's native MTP. This effort has largely been carried out by the Internet Engineering Task Force (IETF) SigTran (Signaling Transport) working group. The protocols derived by the SigTran working group so far are outside the scope of this introductory chapter on SS7. However, full details of SigTran can be found in Chapter 14, "SS7 in the Converged World."

TUP and ISUP both perform the signaling required to set up and tear down telephone calls. As such, both are circuit-related signaling protocols. TUP was the first call control protocol specified. It could support only plain old telephone service (POTS) calls. Most countries are replacing TUP with ISUP. Both North America and Japan bypassed TUP and went straight from earlier signaling systems to ISUP. ISUP supports both POTS and ISDN calls. It also has more flexibility and features than TUP.

With reference to the Open System Interconnection (OSI) seven-layer reference model, SS7 uses a four-level protocol stack. OSI Layer 1 through 3 services are provided by the MTP together with the SCCP. The SS7 architecture currently has no protocols that map into OSI

OSI 1..3 => SS7 (MTPs & SCCP)
OSI 4..6 => None
OSI 7 => SS7 (ISUP & TCAP)

Layers 4 through 6. TUP, ISUP, and TCAP are considered as corresponding to OSI Layer 7 [111]. SS7 and the OSI model were created at about the same time. For this reason, they use some differing terminology.

SS7 uses the term *levels* when referring to its architecture. The term *levels* should not be confused with OSI layers, because they do not directly correspond to each other. *Levels* was a term introduced to help in the discussion and presentation of the SS7 protocol stack. Levels 1, 2, and 3 correspond to MTP 1, 2, and 3, respectively. Level 4 refers to an MTP *user*. The term *user* refers to any protocol that directly uses the transport capability provided by the MTP—namely, TUP, ISUP, and SCCP in traditional SS7. The four-level terminology originated back when SS7 had only a call control protocol (TUP) and the MTP, before SCCP and TCAP were added.

The following sections provide a brief outline of protocols found in the introductory SS7 protocol stack, as illustrated in Figure 4-18.

MTP

MTP levels 1 through 3 are collectively referred to as the MTP. The MTP comprises the functions to transport information from one SP to another.

The MTP transfers the signaling message, in the correct sequence, without loss or duplication, between the SPs that make up the SS7 network. The MTP provides reliable transfer and delivery of signaling messages. The MTP was originally designed to transfer circuit-related signaling because no noncircuit-related protocol was defined at the time.

The recommendations refer to MTP1, MTP2, and MTP3 as the physical layer, data link layer, and network layer, respectively. The following sections discuss MTP2 and MTP3. (MTP1 isn't discussed because it refers to the physical network.) For information on the physical aspects of the Public Switched Telephone Network (PSTN), see Chapter 5, "The Public Switched Telephone Network (PSTN)."

MTP2

Signaling links are provided by the combination of MTP1 and MTP2. MTP2 ensures reliable transfer of signaling messages. It encapsulates signaling messages into variable-length SS7 packets. SS7 packets are called signal units (SUs). MTP2 provides delineation of SUs, alignment of SUs, signaling link error monitoring, error correction by retransmission, and flow control. The MTP2 protocol is specific to narrowband links (56 or 64 kbps).

MTP3

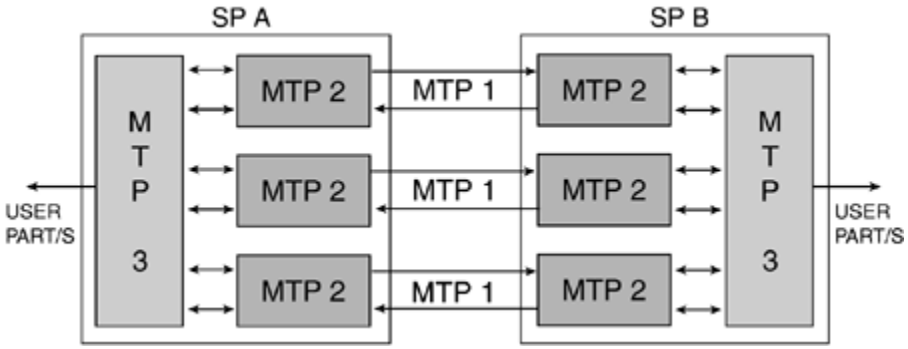
MTP3 performs two functions:

- **Signaling Message Handling (SMH)**— Delivers incoming messages to their intended User Part and routes outgoing messages toward their destination. MTP3 uses the PC to identify the correct node for message delivery. Each message has both an Origination Point Code (OPC) and a DPC. The OPC is inserted into messages at the MTP3 level to identify the SP that originated the message. The DPC is inserted to identify the address of the destination SP. Routing tables within an SS7 node are used to route messages.

- **Signaling Network Management (SNM)**— Monitors linksets and routesets, providing status to network nodes so that traffic can be rerouted when necessary. SNM also provides procedures to take corrective action when failures occur, providing a self-healing mechanism for the SS7 network.

Figure 4-19 shows the relationship between levels 1, 2, and 3.

Figure 4-19. A Single MTP3 Controls Many MTP2s, Each of Which Is Connected to a Single MTP1



TUP and ISUP

TUP and ISUP sit on top of MTP to provide circuit-related signaling to set up, maintain, and tear down calls. TUP has been replaced in most countries because it supports only POTS calls. Its successor, ISUP, supports both POTS and ISDN calls as well as a host of other features and added flexibility. Both TUP and ISUP are used to perform interswitch call signaling. ISUP also has inherent support for supplementary services, such as automatic callback, calling line identification, and so on.

SCCP

The combination of the MTP and the SCCP is called the *Network Service Part (NSP)* in the specifications (but outside the specifications, this term is seldom used).

The addition of the SCCP provides a more flexible means of routing and provides mechanisms to transfer data over the SS7 network. Such additional features are used to support noncircuit-related signaling, which is mostly used to interact with databases (SCPs). It is also used to connect the radio-related components in cellular networks and for inter-SSP communication supporting CLASS services. SCCP also provides application management functions. Applications are mostly SCP database driven and are called subsystems. For example, in cellular networks, SCCP transfers queries and responses between the Visitor Location Register (VLR) and Home Location Register (HLR) databases. Such transfers take place for a number of reasons. The primary reason is to update the subscriber's HLR with the current VLR serving area so that incoming calls can be delivered.

Enhanced routing is called global title (GT) routing. It keeps SPs from having overly large routing tables that would be difficult to provision and maintain. A GT is a directory number that serves as an alias for a physical network address. A physical address consists of a point code and an application reference called a subsystem number (SSN). GT routing allows SPs to use alias addressing to save them from having to maintain overly large physical address tables. Centralized STPs are then used to convert the GT address into a physical address; this process is called Global Title Translation (GTT). This provides the mapping of traditional telephony addresses (phone numbers) to SS7 addresses (PC and/or SSN) for enhanced services. GTT is typically performed at STPs.

NOTE

It is important not to confuse the mapping of telephony numbers using GTT with the translation of telephony numbers done during normal call setup. Voice switches internally map telephony addresses to SS7 addresses during normal call processing using number translation tables. This process does not use GTT. GTT is used only for noncircuit-related information, such as network supplementary services (Calling Name Delivery) or database services (toll-free).

In addition to mapping telephony addresses to SS7 addresses, SCCP provides a set of subsystem management functions to monitor and respond to the condition of subsystems. These management functions are discussed further, along with the other aspects of SCCP, in Chapter 9, "Signaling Connection Control Part (SCCP)."

TCAP

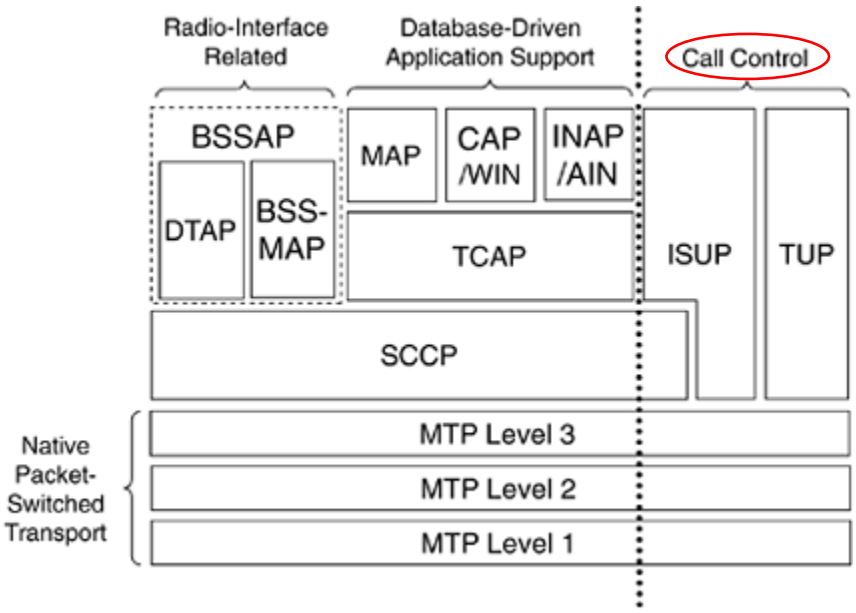
TCAP allows applications (called subsystems) to communicate with each other (over the SS7 network) using agreed-upon data elements. These data elements are called *components*. Components can be viewed as instructions sent between applications. For example, when a subscriber changes VLR location in a global system for mobile communication (GSM) cellular network, his or her HLR is updated with the new VLR location by means of an UpdateLocation component. TCAP also provides transaction management, allowing multiple messages to be associated with a particular communications exchange, known as a transaction.

There are a number of subsystems; the most common are

- Toll-free (E800)
- Advanced Intelligent Network (AIN)
- Intelligent Network Application Protocol (INAP)
- Customizable Applications for Mobile Enhanced Logic (CAMEL)
- Mobile Application Part (MAP)

Figure 4-20 illustrates these subsystems as well as another protocol that uses SCCP, the Base Station Subsystem Application Part. It is used to control the radio-related component in cellular networks.

Figure 4-20. Some Protocols That Might Exist on Top of the SCCP, Depending on the Application



It is highly unlikely that a protocol such as the one shown in Figure 4-20 would exist at any one SP. Instead, protocol stacks vary as required by SP type. For example, because an STP is a routing device, it has only MTP1, MTP2, MTP3, and SCCP. A fixed-line switch without ~~IN~~ support might have only MTP1, MTP2, MTP3, and ISUP, and so forth. A diagram showing how the SS7 protocol stack varies by SP can be found in Chapter 13

Summary

SS7 is a data communications network that acts as the nervous system to bring the components of telecommunications networks to life. It acts as a platform for various services described throughout this book. SS7 nodes are called signaling points (SPs), of which there are three types:

- Service Switching Point (SSP)
- Service Control Point (SCP)
- Signal Transfer Point (STP)

SSPs provide the SS7 functionality of a switch. STPs may be either standalone or integrated STPs (SSP and STP) and are used to transfer signaling messages. SCPs interface the SS7 network to query telecommunication databases, allowing service logic and additional routing information to be obtained to execute services.

SPs are connected to each other using signaling links. Signaling links are logically grouped into a linkset. Links may be referenced as A through F links, depending on where they are in the network.

Signaling is transferred using the packet-switching facilities afforded by SS7. These packets are called signal units (SUs). The Message Transfer Part (MTP) and the Signaling Connection Control Part (SCCP) provide the transfer protocols. MTP is used to reliably transport messages between nodes, and SCCP is used for noncircuit-related signaling (typically, transactions with SCPs). The ISDN User Part (ISUP) is used to set up and tear down both

ordinary (analog subscriber) and ISDN calls. The Transaction Capabilities Application Part (TCAP) allows applications to communicate with each other using agreed-upon data components and manages transactions

Chapter 5. The Public Switched Telephone Network (PSTN)

The term Public Switched Telephone Network (PSTN) describes the various equipment and interconnecting facilities that provide phone service to the public. The network continues to evolve with the introduction of new technologies. The PSTN began in the United States in 1878 with a manual mechanical switchboard that connected different parties and allowed them to carry on a conversation. Today, the PSTN is a network of computers and other electronic equipment that converts speech into digital data and provides a multitude of sophisticated phone features, data services, and mobile wireless access.

TIP

PSTN voice facilities transport speech or voice-band data (such as fax/modems and digital data), which is data that has been modulated to voice frequencies.

At the core of the PSTN are digital switches. The term "switch" describes the ability to cross-connect a phone line with many other phone lines and switching from one connection to another. The PSTN is well known for providing reliable communications to its subscribers. The phrase "five nines reliability," representing network availability of 99.999 percent for PSTN equipment, has become ubiquitous within the telecommunications industry.

This chapter provides a fundamental view of how the PSTN works, particularly in the areas of signaling and digital switching. SS7 provides control signaling for the PSTN, so you should understand the PSTN infrastructure to fully appreciate how it affects signaling and switching. This chapter is divided into the following sections:

- Network Topology
- PSTN Hierarchy
- Access and Transmission Facilities
- Network Timing
- The Central Office
- Integration of SS7 into the PSTN
- Evolving the PSTN to the Next Generation

We conclude with a summary of the PTSN infrastructure and its continuing evolution.

Network Topology

The topology of a network describes the various network nodes and how they interconnect. Regulatory policies play a major role in exactly how voice network topologies are defined in each country, but general similarities exist. While topologies in competitive markets

represent an interconnection of networks owned by different service providers, monopolistic markets are generally an interconnection of switches owned by the same operator.

Depending on geographical region, PSTN nodes are sometimes referred to by different names. The three node types we discuss in this chapter include:

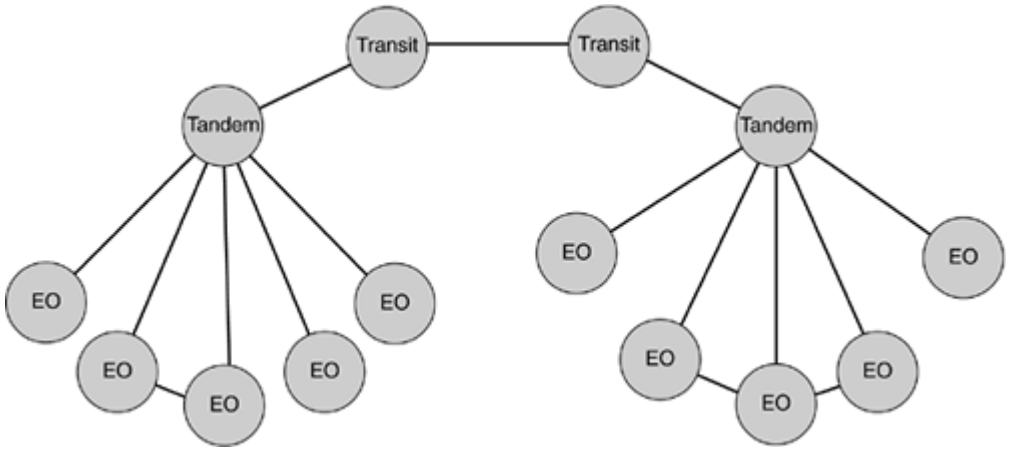
- **End Office (EO)**— Also called a Local Exchange. The End Office provides network access for the subscriber. It is located at the bottom of the network hierarchy.
- **Tandem**— Connects EOs together, providing an aggregation point for traffic between them. In some cases, the Tandem node provides the EO access to the next hierarchical level of the network.
- **Transit**— Provides an interface to another hierarchical network level. Transit switches are generally used to aggregate traffic that is carried across long geographical distances.

There are two primary methods of connecting switching nodes. The first approach is a mesh topology, in which all nodes are interconnected. This approach does not scale well when you must connect a large number of nodes. You must connect each new node to every existing node. This approach does have its merits, however; it simplifies routing traffic between nodes and avoids bottlenecks by involving only those switches that are in direct communication with each other. The second approach is a hierarchical tree in which nodes are aggregated as the hierarchy traverses from the subscriber access points to the top of the tree. PSTN networks use a combination of these two methods, which are largely driven by cost and the traffic patterns between exchanges.

Figure 5-1 shows a generic PSTN hierarchy, in which End Offices are connected locally and through tandem switches. Transit switches provide further aggregation points for connecting multiple tandems between different networks. While actual network topologies vary, most follow some variation of this basic pattern.

Figure 5-1. Generic PSTN Hierarchies

[View full size image]



PSTN Hierarchy

The PSTN hierarchy is implemented differently in the United States and the United Kingdom. The following sections provide an overview of the PSTN hierarchy and its related terminology in each of these countries.

PSTN Hierarchy in the United States

In the United States, the PSTN is generally divided into three categories:

- Local Exchange Networks
- InterExchange Networks
- International Networks

Local Exchange Carriers (LECs) operate Local Exchange networks, while InterExchange Carriers (IXCs) operate InterExchange and International networks.

The PSTN hierarchy in the United States is also influenced by market deregulation, which has allowed service providers to compete for business and by the divestiture of Bell.

Local Exchange Network

The Local Exchange network consists of the digital switching nodes (EOs) that provide network access to the subscriber. The Local Exchange terminates both lines and trunks, providing the subscriber access to the PSTN.

A Tandem Office often connects End Offices within a local area, but they can also be connected directly. In the United States, Tandem Offices are usually designated as either Local Tandem (LT) or Access Tandem (AT). The primary purpose of a Local Tandem is to provide interconnection between End Offices in a localized geographic region. An Access Tandem provides interconnection between local End Offices and serves as a primary point of access for IXCs. Trunks are the facilities that connect all of the offices, thereby transporting inter-nodal traffic.

InterExchange Network

The InterExchange network is comprised of digital switching nodes that provide the connection between Local Exchange networks. Because they are points of high traffic aggregation and they cover larger geographical distances, high-speed transports are typically used between transit switches. In the deregulated U.S. market, transit switches are usually referred to as *carrier switches*. In the U.S., IXCs access the Local Exchange network at designated points, referred to as a Point of Presence (POP). POPs can be connections at the Access Tandem, or direct connections to the End Office.

International Network

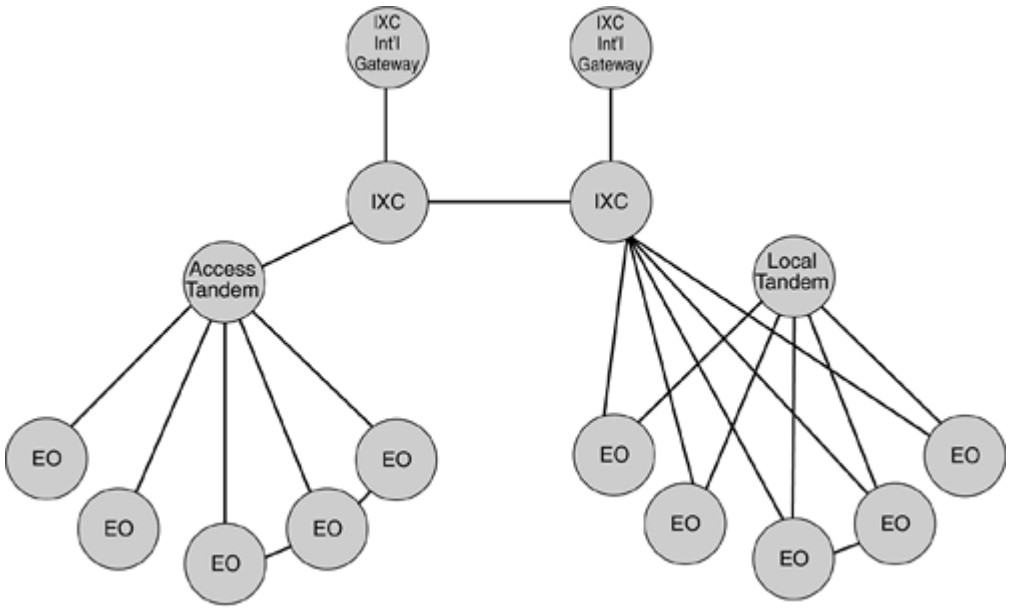
The International network consists of digital switching nodes, which are located in each country and act as international gateways to destinations outside of their respective countries. These gateways adhere to the ITU international standards to ensure interoperability between national networks. The international switch also performs the protocol conversions between national and international signaling. The gateway also performs PCM conversions between A-law and μ-law to produce compatible speech encoding between networks, when necessary.

Service Providers

Deregulation policies in the United States have allowed network operators to compete for business, first in the long-distance market (InterExchange and International) beginning in the mid 1980s, and later in the local market in the mid 1990s. As previously mentioned, LECs operate Local Exchange networks, while IXCs operate the long-distance networks. Figure 5-2 shows a typical arrangement of LEC-owned EOs and tandems interconnected to IXC-owned transit switches. The IXC switches provide long-haul transport between Local Exchange networks, and international connections through International gateway switches.

Figure 5-2. Generic U.S. Hierarchies

[View full size image]



Over the last several years, the terms ILEC and CLEC have emerged within the Local Exchange market to differentiate between the Incumbent LECs (ILECS) and the Competitive LECs (CLECS). ILECs are the incumbents, who own existing access lines to residences and corporate facilities; CLECs are new entrants into the Local Exchange market. Most of the ILECs in the United States came about with the divestiture of AT&T into the seven Regional Bell Operating Companies (RBOC). The remainder belonged to Independent Operating Companies (IOCs). Most of these post-divestiture companies have been significantly transformed today by mergers and acquisitions in the competitive market. New companies have experienced difficulty entering into the Local Exchange market, which is dominated by ILECs. The ILECs own the wire to the subscriber's home, often called the "last mile" wiring. Last mile wiring is expensive to install and gives the ILECs tremendous market leverage. The long-distance market has been easier for new entrants because it does not require an investment in last mile wiring.

Pre-Divestiture Bell System Hierarchy

Vestiges of terminology relating to network topology remain in use today from the North American Bell System's hierarchy, as it existed prior to divestiture in 1984. Telephone switching offices are often still referred to by *class*. For example, an EO is commonly called a class 5 office, and an AT is called a class 4 office. Before divestiture, each layer of the network hierarchy was assigned a class number.

Prior to divestiture, offices were categorized by class number, with class 1 being the highest office category and class 5 being the lowest (nearest to subscriber access). Aggregation of transit phone traffic moved from the class 5 office up through the class 1 office. Each class of traffic aggregation points contained a smaller number of offices. Table 5-1 lists the class categories and office types used in the Bell System Hierarchy.

Table 5-1. Pre-Divestiture Class Categories and Office Types

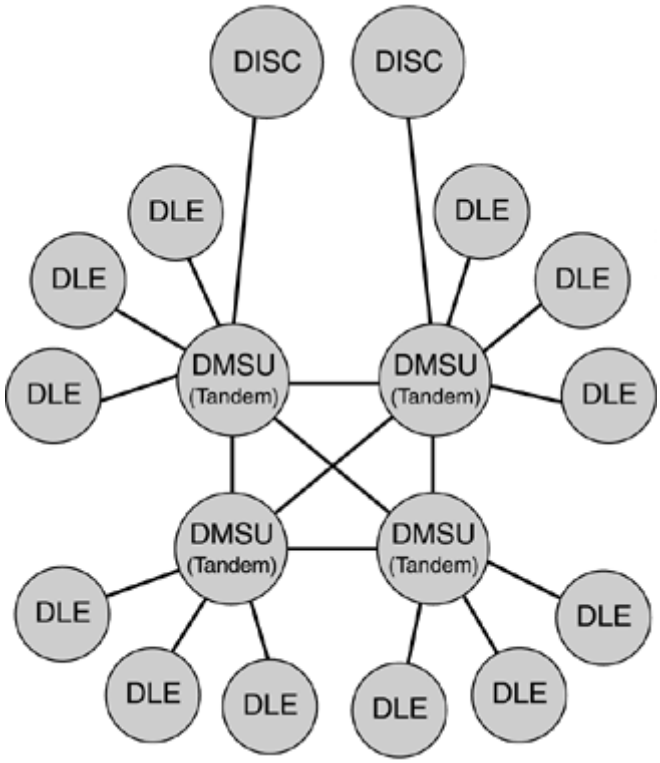
Class	Office Type
1	Regional Center
2	Sectional Center
3	Primary Center
4	Toll Center
5	End Office

Local calls remained within class 5 offices, while a cross-country call traversed the hierarchy up to a regional switching center. This system no longer exists, but we included it to give relevance to the class terminology, which the industry still uses often.

PSTN Hierarchy in the United Kingdom

Figure 5-3 shows the PSTN topology used in the United Kingdom. End Offices are referred to as Digital Local Exchanges (DLE). A fully meshed tandem network of Digital Main Switching Units (DMSU) connects the DLEs. Digital International Switching Centers (DISC) connect the DMSU tandem switches for international call connections.

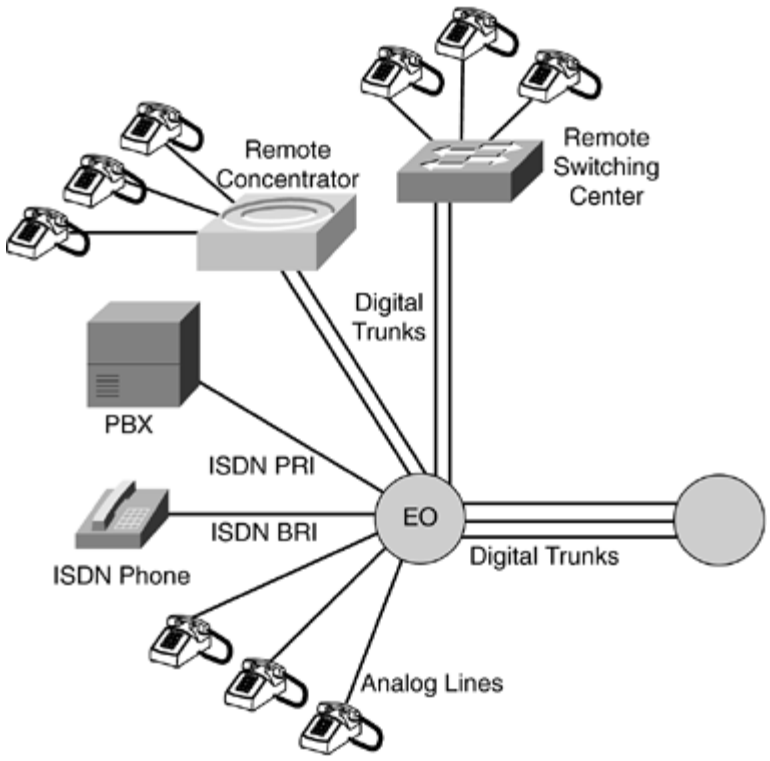
Figure 5-3. U.K. PSTN Hierarchy



Access and Transmission Facilities

Connections to PSTN switches can be divided into two basic categories: lines and trunks. Individual telephone lines connect subscribers to the Central Office (CO) by wire pairs, while trunks are used to interconnect PSTN switches. Trunks also provide access to corporate phone environments, which often use a Private Branch eXchange (PBX)—or in the case of some very large businesses, their own digital switch. Figure 5-4 illustrates a number of common interfaces to the Central Office.

Figure 5-4. End Office Facility Interfaces



Lines

Lines are used to connect the subscriber to the CO, providing the subscriber access into the PSTN. The following sections describe the facilities used for lines, and the access signaling between the subscriber and the CO.

- The Local Loop
- Analog Line Signaling
- Dialing
- Ringing and Answer
- Voice Encoding
- ISDN BRI

The Local Loop

The local loop consists of a pair of copper wires extending from the CO to a residence or business that connects to the phone, fax, modem, or other telephony device. The wire pair consists of a tip wire and a ring wire. The terms *tip* and *ring* are vestiges of the manual switchboards that were used a number of years ago; they refer to the tip and ring of the actual switchboard plug operators used to connect calls. The local loop allows a subscriber to access the PSTN through its connection to the CO. The local loop terminates on the Main Distribution Frame (MDF) at the CO, or on a remote line concentrator.

Remote line concentrators, also referred to as Subscriber Line Multiplexers or Subscriber Line Concentrators, extend the line interface from the CO toward the subscribers, thereby reducing the amount of wire pairs back to the CO and converting the signal from analog to

digital closer to the subscriber access point. In some cases, remote switching centers are used instead of remote concentrators.

Remote switching centers provide local switching between subtending lines without using the resources of the CO. *Remotes,* as they are often generically referred to, are typically used for subscribers who are located far away from the CO. While terminating the physical loop, remotes transport the digitized voice stream back to the CO over a trunk circuit, in digital form.

Analog Line Signaling

Currently, most phone lines are analog phone lines. They are referred to as analog lines because they use an analog signal over the local loop, between the phone and the CO. The analog signal carries two components that comprise the communication between the phone and the CO: the voice component, and the signaling component.

The signaling that takes place between the analog phone and the CO is called in-band signaling. In-band signaling is primitive when compared to the out-of-band signaling used in access methods such as ISDN; see the "ISDN BRI" section in this chapter for more information. DC current from the CO powers the local loop between the phone and the CO. The voltage levels vary between different countries, but an on-hook voltage of –48 to –54 volts is common in North America and a number of other geographic regions, including the United Kingdom.

TIP

The actual line loop voltage varies, based on the distance and the charge level of the batteries connected to the loop at the CO. When the phone receiver is on-hook, the CO sees practically no current over the loop to the phone set. When the phone is off-hook, the resistance level changes, changing the current seen at the CO. The actual amount of loop current that triggers an on/off-hook signal also varies among different countries. In North America, a current flow of greater than 20 milliamps indicates an off-hook condition. When the CO has detected the off-hook condition, it provides a dial tone by connecting a tone generation circuit to the line.

Dialing

When a subscriber dials a number, the number is signaled to the CO as either a series of pulses based on the number dialed, or by Dual Tone Multi-Frequency (DTMF) signals. The DTMF signal is a combination of two tones that are generated at different frequencies. A total of seven frequencies are combined to provide unique DTMF signals for the 12 keys (three columns by four rows) on the standard phone keypad. Usually, the dialing plan of the CO determines when all digits have been collected.

Ringing and Answer

To notify the called party of an incoming call, the CO sends AC ringing voltage over the local loop to the terminating line. The incoming voltage activates the ringing circuit within the phone to generate an audible ring signal. The CO also sends an audible ring-back tone over the originating local loop to indicate that the call is proceeding and the destination phone is ringing. When the destination phone is taken off-hook, the CO detects the change in loop

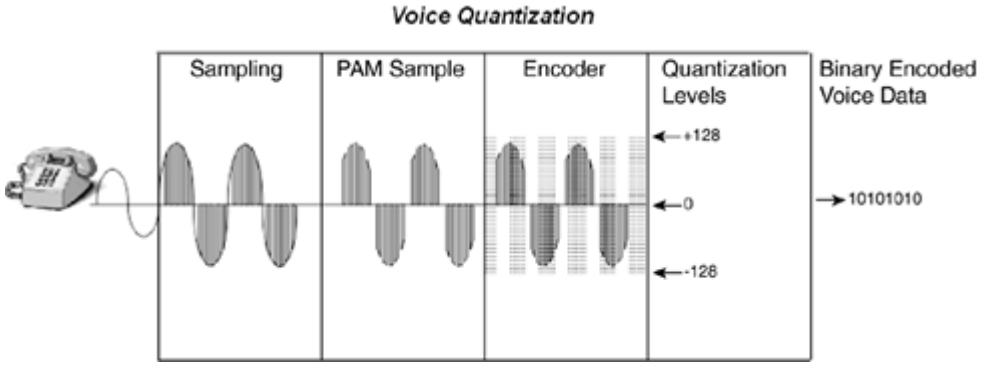
current and stops generating the ringing voltage. This procedure is commonly referred to as *ring trip*. The off-hook signals the CO that the call has been answered; the conversation path is then completed between the two parties and other actions, such as billing, can be initiated, if necessary.

Voice Encoding

An analog voice signal must be encoded into digital information for transmission over the digital switching network. The conversion is completed using a codec (coder/decoder), which converts between analog and digital data. The ITU G.711 standard specifies the Pulse Coded Modulation (PCM) method used throughout most of the PSTN. An analog-to-digital converter samples the analog voice 8000 times per second and then assigns a quantization value based on 256 decision levels. The quantization value is then encoded into a binary number to represent the individual data point of the sample. Figure 5-5 illustrates the process of sampling and encoding the analog voice data.

Figure 5-5. Voice Encoding Process

[View full size image]



Two variations of encoding schemes are used for the actual quantization values: A-law and μ-Law encoding. North America uses μ-Law encoding, and European countries use A-law encoding. When voice is transmitted from the digital switch over the analog loop, the digital voice data is decoded and converted back into an analog signal before transmitting over the loop.

The emergence of voice over IP (VoIP) has prompted the use of other voice-encoding standards, such as ITU G.723, G.726, and ITU G.729. These encoding methods use algorithms that produce more efficient and compressed data, making them more suitable for use in packet networks. Each encoding method involves trade-offs between bandwidth, processing power required for the encoding/decoding function, and voice quality. For example, G.711 encoding/decoding requires little processing and produces high quality speech, but consumes more bandwidth. In contrast, G.723.1 consumes little bandwidth, but requires more processing power and results in lower quality speech.

ISDN BRI

Although Integrated Services Digital Network (ISDN) deployment began in the 1980s, it has been a relatively slow-moving technology in terms of number of installations. ISDN moves

the point of digital encoding to the customer premises. Combining ISDN on the access portion of the network with digital trunks on the core network provides total end-to-end digital connectivity. ISDN also provides out-of-band signaling over the local loop. ISDN access signaling coupled with SS7 signaling in the core network achieves end-to-end out-of-band signaling. ISDN access signaling is designed to complement SS7 signaling in the core network.

There are two ISDN interface types: Basic Rate Interface (BRI) for lines, and Primary Rate Interface (PRI) for trunks. BRI multiplexes two bearer (2B) channels and one signaling (D) channel over the local loop between the subscriber and the CO; this is commonly referred to as 2B+D. The two B channels each operate at 64 kb/s and can be used for voice or data communication. The D channel operates at 16 kb/s and is used for call control signaling for the two B channels. The D channel can also be used for very low speed data transmission. Within the context of ISDN reference points, the local loop is referred to as the U-loop. It uses different electrical characteristics than those of an analog loop.

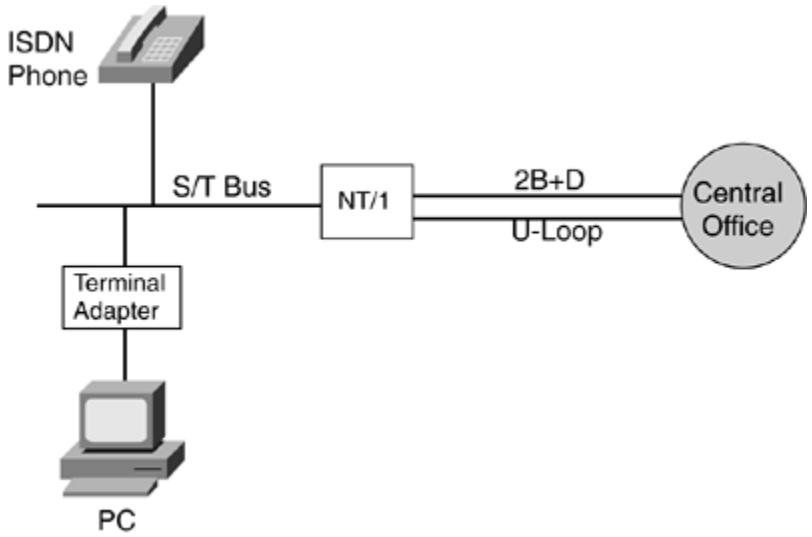
Voice quantization is performed within the ISDN phone (or a Terminal Adapter, if an analog phone is used) and sent to a local bus: the S/T bus. The S/T bus is a four-wire bus that connects local ISDN devices at the customer premises to a Network Termination 1 (NT1) device. The NT1 provides the interface between the Customer Premises Equipment (CPE) and the U-loop.

TIP

CPE refers to any of the ISDN-capable devices that are attached to the S/T bus.

The NT1 provides the proper termination for the local S/T bus to individual devices and multiplexes the digital information from the devices into the 2B+D format for transmission over the U-loop. Figure 5-6 illustrates the BRI interface to the CO. Only ISDN devices connect directly to the S/T bus. The PC uses an ISDN Terminal Adapter (TA) card to provide the proper interface to the bus.

Figure 5-6. ISDN Basic Rate Interface



The ISDN U-Loop terminates at the CO on a line card that is specifically designed to handle the 2B+D transmission format. The call control signaling messages from the D channel are designed to map to SS7 messages easily for outbound calls over SS7 signaled trunks.

TIP

For U.S. networks, the Telcordia TR-444 (Generic Switching Systems Requirements Supporting ISDN access using the ISDN User Part) standard specifies the inter-working of ISDN and SS7.

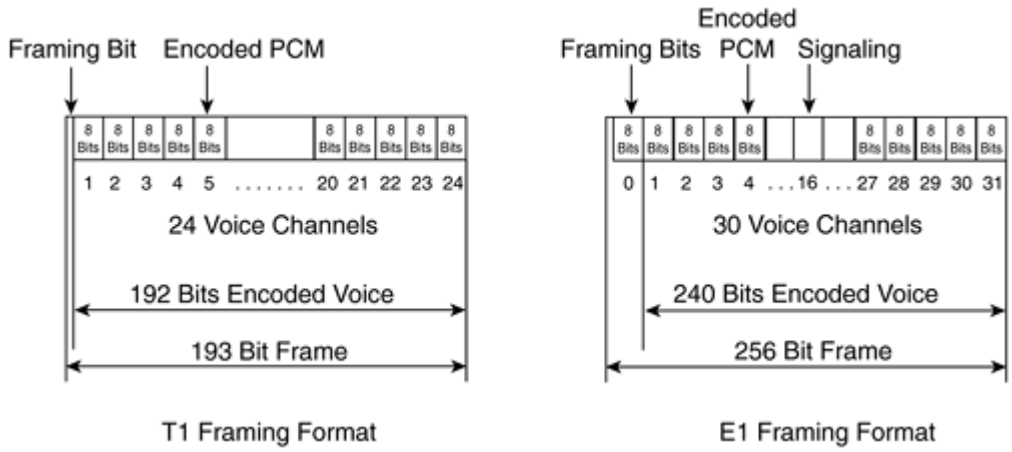
Trunks

Trunks carry traffic between telephony switching nodes. While analog trunks still exist, most trunks in use today are digital trunks, which are the focus of this section. Digital trunks may be either four-wire (twisted pairs) or fiber optic medium for higher capacity. T1 and E1 are the most common trunk types for connecting to End Offices. North American networks use T1, and European networks use E1.

On the T1/E1 facility, voice channels are multiplexed into digital bit streams using Time Division Multiplexing (TDM). TDM allocates one timeslot from each digital data stream's frame to transmit a voice sample from a conversation. Each frame carries a total of 24 multiplexed voice channels for T1 and 30 channels for E1. The T1 frame uses a single bit for framing, while E1 uses a byte. Figure 5-7 shows the formats for T1 and E1 framing.

Figure 5-7. T1/E1 Framing Formats

[View full size image]



The E1 format also contains a channel dedicated to signaling when using in-band signaling. The T1 format uses "robbed bit" signaling when using in-band signaling. The term "robbed bit" comes from the fact that bits are taken from the PCM data to convey trunk supervisory signals, such as on/off-hook status and winks. This is also referred to as A/B bit signaling. In every sixth frame, the least significant bits from each PCM sample are used as signaling bits. In the case of Extended Superframe trunks (ESF), A/B/C/D bits are used to indicate trunk supervision signals. A/B bit signaling has been widely replaced by SS7 signaling, but it still exists in some areas.

Trunks are multiplexed onto higher capacity transport facilities as traffic is aggregated toward tandems and transit switches. The higher up in the switching hierarchy, the more likely optical fiber will be used for trunk facilities for its increased bandwidth capacity. In North America, Synchronous Optical Network (SONET) is the standard specification for transmission over optical fiber. SONET defines the physical interface, frame format, optical line rates, and an OAM&P protocol. In countries outside of North America, Synchronous Digital Hierarchy (SDH) is the equivalent optical standard. Fiber can accommodate a much higher bandwidth than copper transmission facilities, making it the medium of choice for high-density trunking.

Standard designations describe trunk bandwidth in terms of its capacity in bits/second. The basic unit of transmission is Digital Signal 0 (DS0), representing a single 64 kb/s channel that occupies one timeslot of a Time Division Multiplex (TDM) trunk. Transmission rates are calculated in multiples of DS0 rates. For example, a T1 uses 24 voice channels at 64 kb/s per channel to produce a DS1 transmission rate of 1.544 mb/s, calculated as follows:

24 x 64 kb/s = 1.536 kb/s + 8000 b/s framing bits = 1.544 mb/s

The optical transmission rates in the SONET transport hierarchy are designated in Optical Carrier (OC) units. OC-1 is equivalent to T3. Higher OC units are multiples of OC-1; for example, OC-3 is simply three times the rate of OC-1. In North America, the electrical equivalent signals are designated as Synchronous Transport Signal (STS) levels. The ITU SDH standard uses the STM to designate the hierarchical level of transmission. Table 5-2 summarizes the electrical transmission rates, and Table 5-3 summarizes the SONET/SDH transmission rates.

Table 5-2. Electrical Transmission Rates

Designation	Voice Channels	Transmission Rate mb/s
T1 (North America)	24	1.544
E1 (Europe)	30	2.048
E3 (Europe)	480	34.368
T3 (North America)	672	44.736

Table 5-3. SONET/SDH Transmission Rates

SONET Optical Level	SONET Electrical Level	SDH Level	Voice Channels	Transmission Rate mb/s
OC-1	STS-1	—	672	51.840
OC-3	STS-3	STM-1	2016	155.520
OC-12	STS-12	STM-4	8064	622.080
OC-48	STS-48	STM-16	32,256	2488.320
OC-96	STS-96	STM-32	64,512	4976.64

Table 5-3. SONET/SDH Transmission Rates

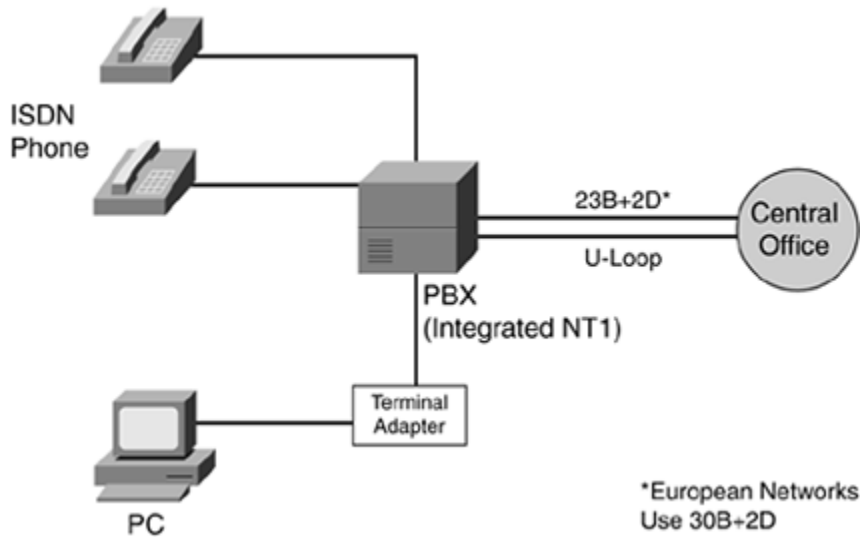
SONET Optical Level	SONET Electrical Level	SDH Level	Voice Channels	Transmission Rate mb/s
OC-192	STS-192	STM-64	129,024	9953.280
OC-768	STS-768	STM-256	516,096	39,813.120

In addition to copper and fiber transmission mediums, microwave stations and satellites are also used to communicate using radio signals between offices. This is particularly useful where it is geographically difficult to install copper and fiber into the ground or across rivers.

ISDN PRI

Primary Rate Interface (PRI) provides ISDN access signaling over trunks and is primarily used to connect PBXs to the CO. As with BRI, PRI converts all data at the customer premises into digital format before transmitting it over the PRI interface. In the United States, PRI uses 23 bearer channels for voice/data and one signaling channel for call control. The single signaling channel handles the signaling for calls on the other 23 channels. This scheme is commonly referred to as 23B+D. Each channel operates at a rate of 64 kb/s. Figure 5-8 illustrates a PBX connected to the CO through a PRI trunk.

Figure 5-8. ISDN Primary Rate Interface



Other variations of this scheme use a single D channel to control more than 23 bearer channels. You can also designate a channel as a backup D channel to provide redundancy in case of a primary D channel failure. In the United States, U-Loop for PRI is a four-wire interface that operates at 1.544 mb/s. The U-Loop terminates to an NT1, which is typically integrated into the PBX at the customer premises.

In Europe, PRI is based on 32 channels at a transmission rate of 2.048 mb/s. There are 30 bearer channels and two signaling channels, which are referred to as 30B+2D.

Network Timing

Digital trunks between two connecting nodes require clock synchronization in order to ensure proper framing of the voice channels. The sending switch clocks the bits in each frame onto the transmission facility. They are clocked into the receiving switch at the other end of the facility. Digital facility interfaces use buffering techniques to store the incoming frame and accommodate slight variation in the timing of the data sent between the two ends. A problem arises if the other digital switch that is connected to the facility has a clock signal that is out of phase with the first switch. The variation in clock signals eventually causes errors in identifying the beginning of a frame. This condition is known as *slip*, and it results in buffer overrun or buffer underrun. Buffer overrun occurs if the frequency of the sending clock is greater than the frequency of the receiving clock, discarding an entire frame of data. Buffer underrun occurs if the frequency of the sending clock is less than the frequency of the receiving clock, repeating a frame of data. Occasional slips do not present a real problem for voice calls, although excessive slips result in degraded speech quality. However, they are more detrimental to the data transfer, in which each bit is important. Therefore, synchronization of time sources between the digital switches is important. Because digital transmission facilities connect switches throughout the network, this requirement escalates to a network level, where the synchronization of many switches is required.

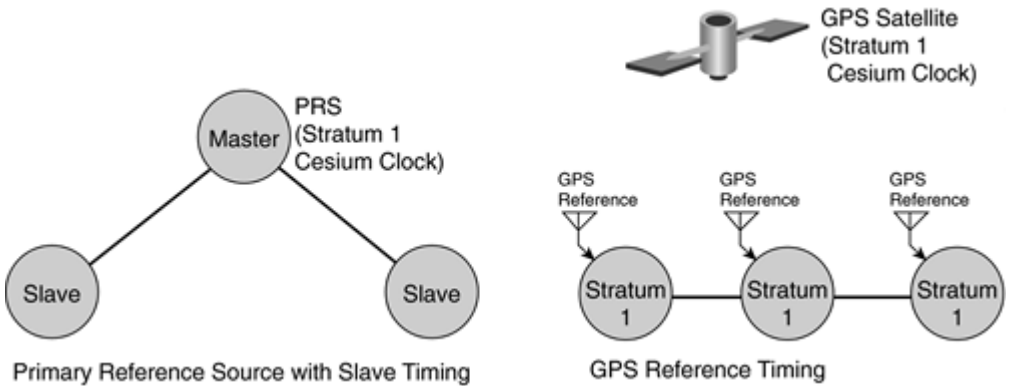
There are various methods of synchronizing nodes. One method involves a single master clock source, from which other nodes derive timing in a master/slave arrangement. Another method uses a plesiochronous arrangement, where each node contains an independent clock whose accuracy is so great that it remains independently synchronized with other nodes. You can also use a combination of the two methods by using highly accurate clocks as a Primary Reference Source (PRS) in a number of nodes, providing timing to subtending nodes in the network.

The clocks' accuracy is rated in terms of stratum levels. Stratums 1 through 4 denote timing sources in order of descending accuracy. A stratum 1 clock provides the most accurate clock source with a free-running accuracy of $\pm 1 \times 10^{-11}$, meaning only one error can occur in 10^{11} parts. A stratum 4 clock provides an accuracy of $\pm 32 \times 10^{-6}$.

Since the deployment of Global Positioning System (GPS) satellites, each with a number of atomic clocks on-board, GPS clocks have become the preferred method of establishing a clock reference signal. Having a GPS clock receiver at each node that receives a stratum 1-quality timing signal from the GPS satellite flattens the distributed timing hierarchy. If the GPS receiver loses the satellite signal, the receiver typically runs free at stratum 2 or less. By using a flattened hierarchy based on GPS receivers, you remove the need to distribute the clock signal and provide a highly accurate reference source for each node. Figure 5-9 shows an example that uses a stratum 1 clock at a digital switching office to distribute timing to subtending nodes, and also shows an example that uses a GPS satellite clock receiver at each office.

Figure 5-9. Network Timing for Digital Transmission

[View full size image]



SS7 links are subject to the same timing constraints as the trunk facilities that carry voice/data information because they use digital trunk transmission facilities for transport. If they produce unrecoverable errors, slips on the transmission facilities might affect SS7 messages. Therefore, you must always consider network timing when establishing SS7 links between nodes in the PSTN

The Central Office

The Central Office (CO) houses the digital switching equipment that terminates subscribers' lines and trunks and *switch* calls. The term *switch* is a vestige of the switchboard era, when call connections were manually created using cords to connect lines on a plugboard. Electro-mechanical switches replaced manual switchboards, and those eventually evolved into the computer-driven digital switches of today's network. Now switching between calls is done electronically, under software control.

The following section focuses on these areas of the CO:

- The Main Distribution Frame
- The Digital Switch
- The Switching Matrix
- Call Processing

Main Distribution Frame

Incoming lines and trunks are terminated on the Main Distribution Frame (MDF). The MDF provides a junction point where the external facilities connect to the equipment within the CO. Jumpers make the connections between the external facilities and the CO equipment, thereby allowing connections to be changed easily. Line connections from the MDF to the digital switching equipment terminate on line cards that are designed to interface with the particular type of line being connected—such as POTS, ISDN BRI, and Electronic Key Telephone Set (EKTS) phone lines. For analog lines, this is normally the point at which voice encoding takes place. Trunk connections from the MDF are terminated on trunk interface cards, providing the necessary functions for message framing, transmission, and reception.

The Digital Switch

The digital switch provides a software-controlled matrix of interconnections between phone subscribers. A handful of telecommunications vendors produce the digital switches that comprise the majority of the modern PSTN; Nortel, Lucent, Siemens, Alcatel, and Ericsson hold the leading market share. While the digital switch's basic functionality is common across vendors, the actual implementation is vendor dependent. This section provides a general perspective on the functions of the digital switch that are common across different implementations.

All digital switches are designed with some degree of distributed processing. A typical architecture includes a central processing unit that controls peripheral processors interfacing with the voice channels. Redundancy is always employed in the design to provide the high reliability that is expected in the telephony network. For example, the failure of one central processing unit results in the activation of an alternate processing unit.

The line and trunk interface cards, mentioned previously, represent the point of entry into the digital switch. These cards typically reside in peripheral equipment that is ultimately controlled by the central processor. Within the digital switch, all voice streams are digitized data. Some voice streams, such as those from ISDN facilities and digital trunks, enter the switch as digital data. Other voice streams, such as the analog phone, enter as analog data but undergo digital conversion at their point of entry. Analog lines interface with line cards that contain codecs, which perform the PCM processing to provide digital data to the switch and analog data to the line. Using the distributed processing architecture, many functions related to the individual voice channels are delegated to the peripheral interface equipment. This relieves the central processor of CPU intensive, low-level processing functions, such as scanning for on/off hooks on each individual line to determine when a subscriber wants to place a call.

The central processing unit monitors information from peripheral processors on call events—such as origination, digit collection, answer, and termination—and orchestrates the actual call setup and release. Information from these events is also used to perform call accounting, billing, and statistical information such as Operational Measurements (OM).

Although the main purpose of the digital switch is to perform call processing, much of its functionality is dedicated to maintenance, diagnostics, and fault recovery to ensure reliability.

TIP

An OM is a counter that records an event of particular interest, such as the number of call attempts or the number of a particular type of message received, to service providers. OMs can also be used to record usage in terms of how long a resource is used. Modern digital switches usually record hundreds, or even thousands of different types of OMs for various events taking place in the switch.

Switching Matrix

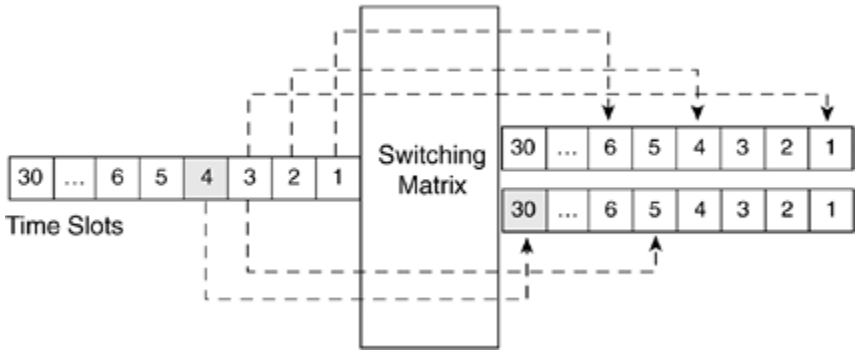
A modern digital switch can process many voice channels. The actual number of channels it processes varies with the switch vendor and particular model of switch, but they often

process tens of thousands of voice channels in a single switch. A number of switches have capacities of over 100,000 connections.

The switch is responsible for many tasks, but one of its primary functions is connecting voice channels to create a bi-directional conversation path between two phone subscribers. All digital switches incorporate some form of switching matrix to allow the connection of voice channels to other voice channels. Once a circuit is set up between the two subscribers, the connection remains for the duration of the call. This method of setting up call connections is commonly known as circuit switching.

Figure 5-10 illustrates how a switching matrix demultiplexes individual timeslots from a multiplexed stream of voice channels and inserts them into the appropriate time slot for a connection on another facility, to connect voice channels. For example, in the figure, time slot 4 from the digital stream on the left connects to timeslot 30 of the digital stream on the right. The figure shows thirty channels, but the number of channels depends on the individual implementation of the switching matrix.

Figure 5-10. TDM Switching Matrix



Each timeslot represents a voice connection path. The matrix connects the two paths to provide a conversation path between two parties. For long-distance calls that traverse a number of switches, an individual call goes through multiple switching matrices and is mapped to a new timeslot at each switching point. When the call is set up, it occupies the voice channel that was set up through the network for the duration of the call.

Call Processing

Call processing is associated with the setup, maintenance, and release of calls within the digital switch. The process is driven by software, in response to stimulus from the facilities coming into the switch. Signaling indications, such as on/off-hook, dialing digits, and answer, are all part of the stimuli that drive the processing of calls.

Each call process can be represented as an originating call half and a terminating call half. When combined, the two halves are completely representative of the call. The originating half is created when the switch determines that the originator is attempting a call. The terminating call half is created when the destination has been identified, typically at the translations or routing phase. The Intelligent Network standards have established a standardized call model, which incorporates the half-call concept. A complete discussion of the call model is presented in Chapter 11, "Intelligent Networks (IN)."

Call processing can be broken down in various ways; the following list provides a succinct view of the major stages of establishing and disconnecting a call.

- Origination
- Digit Collection
- Translation (Digit Analysis)
- Routing
- Connection
- Disconnection

Additional functions, such as billing and service interactions, also take place, but are excluded in our simple view of processing.

Origination

For a line, this initial phase of call processing occurs when a subscriber goes off-hook to initiate a call. The actual event provided to the digital switch to indicate a line origination can be a change in loop current for analog lines, or a setup message from an ISDN BRI facility. In-band A/B bit off-hook signaling, an ISDN PRI setup message, or an Initial Address Message from an SS7 signaled trunk can signal a digital trunk's origination. All of these events indicate the origination of a new call. The origination event creates the originating half of the call.

Digit Collection

For analog line originations, the switch collects digits as the caller dials them. Inter-digit timing monitors the amount of time the caller takes to dial each digit so that the line cannot be left in the dialing state for an infinite amount of time. If the caller does not supply the required number of digits for calling within a specified time, the caller is usually connected to a digital announcement to indicate that there is a problem with dialing, a Receiver Off-Hook (ROH) tone, or both. The dialing plan used for the incoming facility usually specifies the number of digits that are required for calling.

For ISDN lines, the dialed digits are sent in an ISDN Setup message.

Translation

Translation, commonly referred to as digit analysis, is the process of analyzing the collected digits and mapping them to a result. The translation process directs calls to their network destination. The dial plan associated with the incoming line, or trunk, is consulted to determine how the digits should be translated. Different dial plans can be associated with different incoming facilities to allow flexibility and customization in the translation of incoming calls. The dial plan specifies such information as the minimum and maximum number of digits to be collected, acceptable number ranges, call type, special billing indicators, and so forth. The translation process can be somewhat complex for calls that involve advanced services like Centrex, which is often associated with business phones.

TIP

Centrex is a set of services provided by the local exchange switch to business subscribers, including features like ring again, call parking, and conferencing. Centrex allows businesses

to have many of the services provided by a PBX without the overhead of PBX cost, administration, and maintenance.

The process of digit translation can produce several different results. The most common result is a *route* selection for the call to proceed. Other results include connection to a recorded announcement or tone generator, or the sending of an Intelligent Network Query message for calls involving Intelligent Network services. Network administrators provision dial plan, routing information, and other translation-related information on the switch. However, information returned from IN queries can be used to modify or override statically provisioned information, such as routing.

Routing

The call proceeds to the routing stage after translation processing. Routing is the process of selecting a voice channel (on a facility) over which to send the outbound call toward its intended destination, which the dialed digits identify during translation. Routing typically uses route lists, which contain multiple routes from which to choose. For calls that are destined outside of the switching node, a *trunk group* is selected for the outbound call. A trunk group is a collection of trunk members that are connected to a single destination. After a trunk group is selected, an individual *trunk member* is selected from the group. A trunk member occupies an individual time slot on a digital trunk.

Routing algorithms are generally used for selecting the individual trunk circuit. For example, members of an outgoing trunk group are commonly selected using algorithms such as least idle, most idle, and ascending or descending order (based on the numerical trunk member number).

Connection

Call connection must take place on both the transmit and receive paths for a bi-directional conversation to take place. Each involved switch creates a connection between the originating half of the call and the terminating half of the call. This connection must be made through the switching matrix, and the speech path must be *cut through* between the incoming and outgoing voice channels. Supervision messages or signals sent from the central processor to the peripheral interfaces typically cut through the connection for the speech path. The central processor uses supervision signals to indicate how the peripheral processors should handle lower-level functions. It is typical to cut through the backward speech path (from terminator to originator) before cutting through the forward speech path. This approach allows the terminating switch to send the audible ringback over the voice channel, to the originating switch. When the originating switch receives an answer indication, the call path should be connected in both directions.

Disconnection

A call may be disconnected when it is active, meaning that it has been set up and is in the talking state. Disconnection can be indicated in a several ways. For analog lines, the originating or terminating side of the call can go on-hook, causing a disconnection.

TIP

Actually, the call is not disconnected when the terminating line goes on-hook, in some cases. These cases are examined further in Chapter 8, "ISUP."

ISDN sets send a Disconnect message to disconnect the call. For trunks using in-band signaling, on-hook is signaled using the signaling bits within the voice channel. For SS7 trunks, a Release message is the signal to disconnect a call.

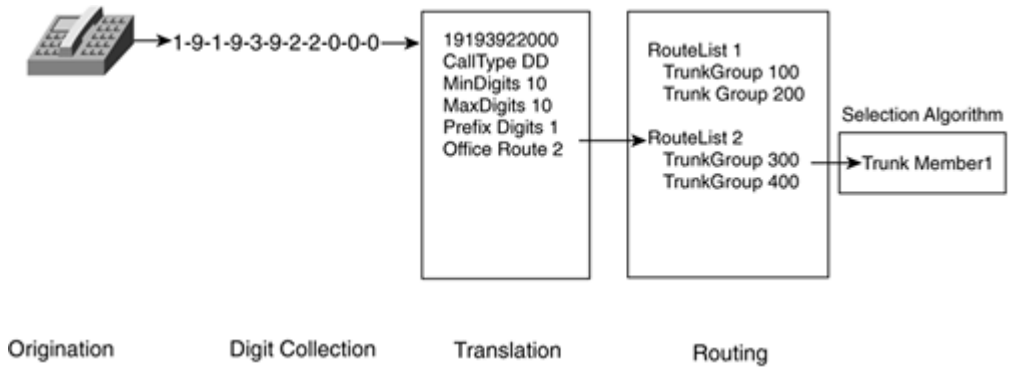
Call Setup

Figure 5-11 shows a typical call setup sequence for a line-to-trunk call. For these calls, the originator dials a number and the digits are collected and processed according to the originating line's dial plan. The dial plan yields a result and points to a list of routes to another switching node. The route list contains a list of trunk groups, from which one group will be selected, usually based on primary and alternate route choices. After the group is selected, an actual trunk member (digital timeslot) is chosen for the outgoing path. The selection of the individual trunk member is typically based on standardized trunk selection algorithms, such as:

- **Most Idle**— The trunk member that has been used the least
- **Least Idle**— The trunk member that has been used the most
- **Ascending**— The next non-busy trunk member, in ascending numerical order
- **Descending**— The next non-busy trunk member, in descending numerical order

Figure 5-11. Basic Origination Call Processing

[View full size image]



Both a call origination endpoint and a call termination endpoint have been established in respect to the digital switch processing the call. The connection can now be made through

the switching matrix between the two endpoints. The timing of the actual speech path cut-through between the external interfaces varies based on many factors, but the switch now has the information it needs to complete the full connection path at the appropriate time, as determined by software

Integration of SS7 into the PSTN

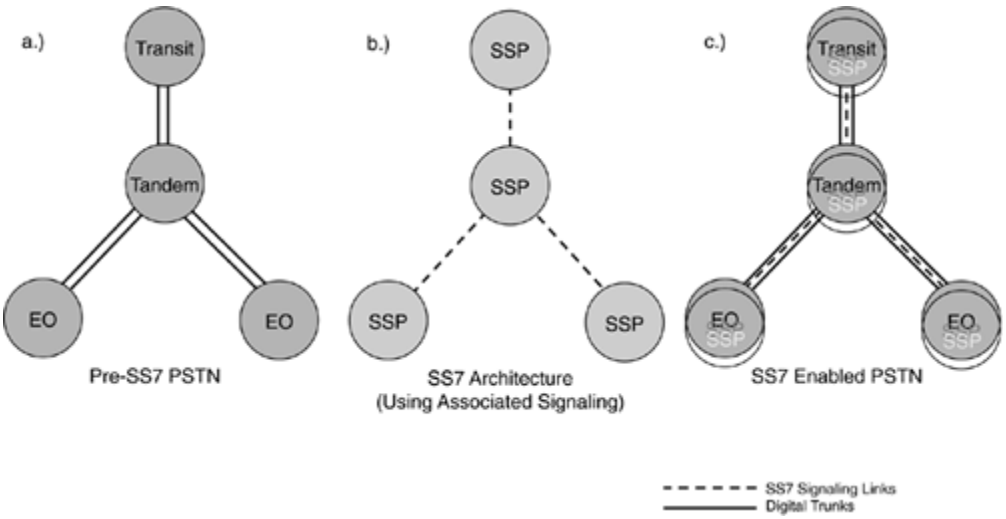
This section provides a brief overview of how the SS7 architecture is applied to the PSTN. Since SS7 has not been presented in great detail, the examples and information are brief and discussed only in the context of the network nodes presented in this section.

The PSTN existed long before SS7. The network's general structure was already in place, and it represented a substantial investment. The performance requirements mandated by the 800 portability act of 1993 was one of the primary drivers for the initial deployment of SS7 by ILECs in the United States. IXCs embraced SS7 early to cut down on post-dial delay which translated into significant savings on access/egress charges. Federal regulation, cost savings, and the opportunity to provide new revenue generating services created a need to deploy SS7 into the existing PSTN.

SS7 was designed to integrate easily into the existing PSTN, to preserve the investment and provide minimal disruption to the network. During SS7's initial deployment, additional hardware was added and digital switches received software upgrades to add SS7 capability to existing PSTN nodes. In the SS7 network, a digital switch with SS7 capabilities is referred to as a Service Switching Point (SSP). When looking at the SS7 network topologies in later chapters, it is important to realize that the SSP is not a new node in the network.

Instead, it describes an existing switching node, to which SS7 capabilities have been added. Similarly, SS7 did not introduce new facilities for signaling links, but used timeslots on existing trunk facilities. PSTN diagrams containing End Offices and tandems connected by trunks represent the same physical facilities as those of SS7 diagrams that show SSP nodes with interconnecting links. The introduction of SS7 added new nodes, such as the STP and SCP; however, all of the switching nodes and facilities that existed before SS7 was introduced are still in place. Figure 5-12 shows a simple view of the PSTN, overlaid with SS7-associated signaling capabilities.

Figure 5-12. SS7 Overlaid onto the PSTN



View *a* in the previous figure shows that trunk facilities provide the path for voice and in-band signaling. View *b* shows the SS7 topology using simple associated signaling for all nodes. View *c* shows the actual SS7-enabled PSTN topology. The existing switching nodes and facilities are enhanced to provide basic SS7 call processing functionality. Although this associated signaling architecture is still quite common in Europe, the United States primarily uses a quasi-associated signaling architecture.

SS7 Link Interface

The most common method for deploying SS7 links is for each link to occupy a timeslot, such as a T1 or E1, on a digital trunk. As shown in Figure 5-12, the signaling links actually travel on the digital trunk transmission medium throughout the network. At each node, the SS7 interface equipment must extract the link timeslot from the digital trunk for processing. This process is typically performed using a channel bank, or a Digital Access and Cross-Connect (DAC), which demultiplexes the TDM timeslot from the digital trunk. The channel bank, or DAC, can extract each of the timeslots from the digital stream, allowing them to be processed individually. The individual SS7 link provides the SS7 messages to the digital switch for processing. While implementations vary, dedicated peripheral processors usually process the lower levels of the SS7 protocol (Level 1, Level 2, and possibly a portion of Level 3); call- and service-related information is passed on to the central processor, or to other peripheral processors that are designed for handling call processing–related messages. Of course, this process varies based on the actual equipment vendor.

Evolving the PSTN to the Next Generation

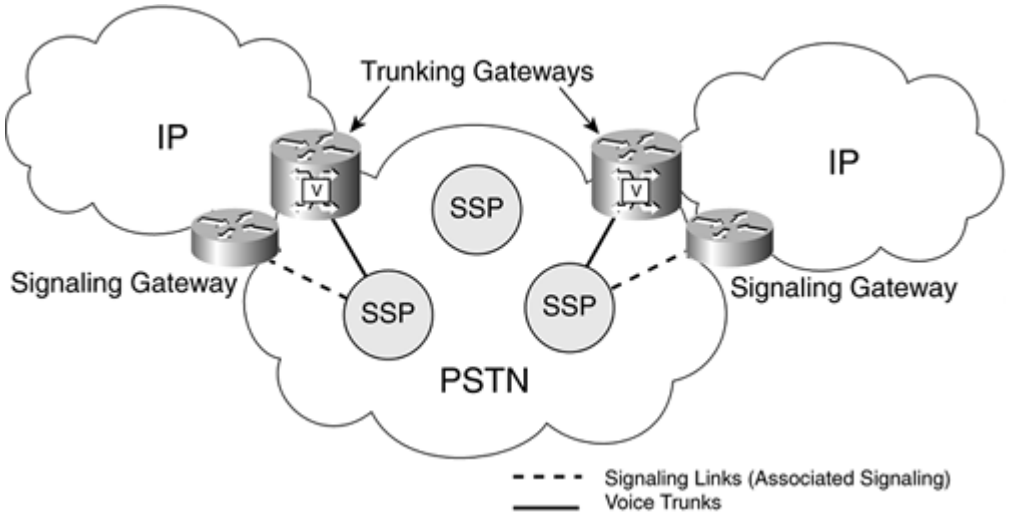
The expansion of the Internet continues to drive multiple changes in the PSTN environment. First, more network capacity is used to transport data over the PSTN. Dial-up Internet services use data connections that are set up over the PSTN to carry voice-band data over circuit-switched connections. This is a much different situation than sending data over a data network. Data networks use packet switching, in which many data transactions share the same facilities. Circuit-switched connections are dedicated connections, which occupy a circuit for the duration of a call. The phone networks were originally engineered for the

three-minute call, which was the average length used for calculations when engineering the voice network. Of course, Internet connections tend to be much more lengthy, meaning that more network capacity is needed. The changes driven by the Internet, however, reach much further than simply an increase in network traffic. Phone traffic is being moved to both private packet-based networks and the public Internet, thereby providing an alternative to sending calls over the PSTN. Several different architectures and protocols are competing in the VoIP market to establish alternatives to the traditional circuit-switched network presented in this chapter. The technologies are not necessarily exclusive; some solutions combine the various technologies. Among the current leading VoIP technologies are:

- Soft switches
- H.323
- Session Initiation Protocol (SIP)

Each of these VoIP architectures use VoIP-PSTN gateways to provide some means of communication between the traditional PSTN networks and VoIP networks. These gateways provide access points for interconnecting the two networks, thereby creating a migration path from PSTN-based phone service to VoIP phone service. The core network interface connections for VoIP into the PSTN are the trunk facilities that carry the voice channels and the signaling links that carry SS7 signaling. PRI is also commonly used for business to network access. Figure 5-13 shows the interconnection of VoIP architectures to the PSTN using signaling gateways and trunking gateways. Chapter 14, "SS7 in the Converged World," discusses these VoIP technologies in more detail.

Figure 5-13. VoIP Gateways to the PSTN



Summary

This chapter provides an overview of the PSTN, as it existed before VoIP technologies emerged. The majority of the PSTN still appears as this chapter presents it. Many of the diagrams in telecommunications literature illustrating *next generation* technologies—such as soft switches, H.323, and Session Initial Protocol (SIP)—show interfaces to the PSTN. The

diagrams refer to the PSTN discussed here, dominated by large, digital switches. The technologies introduced often replace some portion of the existing PSTN; however, they must also remain connected to the existing PSTN to communicate with the rest of the world. The VoIP-PSTN gateways provide this transition point, thus enabling a migration path from the traditional PSTN to the next generation architecture.

While the PSTN varies in its implementation from country to country, a number of common denominators exist. The PSTN is a collection of digital switching nodes that are interconnected by trunks. The network topology is usually a hierarchical structure, but it often incorporates some degree of mesh topology. The topology provides network access to residential and business subscribers for voice and data services. VoIP began another evolution of the PSTN architecture. The PSTN is a large infrastructure that will likely take some time to completely migrate to the next generation of technologies; but this migration process is underway.

Chapter 8. ISDN User Part (ISUP)

The *ISDN User Part (ISUP)* is responsible for setting up and releasing trunks used for inter-exchange calls. As its name implies, ISUP was created to provide core network signaling that is compatible with ISDN access signaling. The combination of ISDN access signaling and ISUP network signaling provides an end-to-end transport mechanism for signaling data between subscribers. Today, the use of ISUP in the network has far exceeded the use of ISDN on the access side. ISUP provides signaling for both non-ISDN and ISDN traffic; in fact, the majority of ISUP-signaled traffic currently originates from analog access signaling, like that used by basic telephone service phones.

The primary benefits of ISUP are its speed, increased signaling bandwidth, and standardization of message exchange. Providing faster call setup times than Channel Associated Signaling (CAS), it ultimately uses trunk resources more effectively. The difference in post-dial delay for calls using ISUP trunks is quite noticeable to the subscriber who makes a call that traverses several switches.

NOTE

Post-dial delay is the time between when the originator dials the last digit and the originating end receives an indication (or audible ringback).

In addition to its speed efficiencies, ISUP enables more call-related information to be exchanged because it uses Common Channel Signaling (CCS). CAS signaling severely limits the amount of information that can be exchanged over trunks because it shares a small amount of space with a call's voice stream. ISUP defines many messages and parameters, therefore, allowing information about a call to be exchanged both within the network and between end-users. Although messages and parameters do vary between different

countries, a given variant provides a standard means of exchanging information between vendor equipment within the national network, and to a large degree, at the international level.

For the reader who is unfamiliar with the PSTN and how switching exchanges work, Chapter 5, "The Public Switched Telephone Network (PSTN)," explains the PSTN, describes the basic concepts of call processing at an exchange, and introduces the concepts of trunks, trunkgroups, and routing.

ISUP consists of call processing, supplementary services, and maintenance functions. This chapter is divided into the following sections, which describe the specific components of ISUP:

- Bearers and Signaling
- ISUP and the SS7 Protocol Stack
- ISUP Message Flow
- Message Timers
- Circuit Identification Codes
- Enbloc and Overlap Address Signaling
- Circuit Glare
- Continuity Test
- ISUP Message Format
- Detailed Call Walk-Through
- Circuit Suspend and Resume
- ISUP and Local Number Portability
- ISUP–ISUP Tandem Calls
- Interworking with ISDN
- Supplementary Services
- Additional Call Processing Messages
- Maintenance Messages and Procedures

Bearers and Signaling

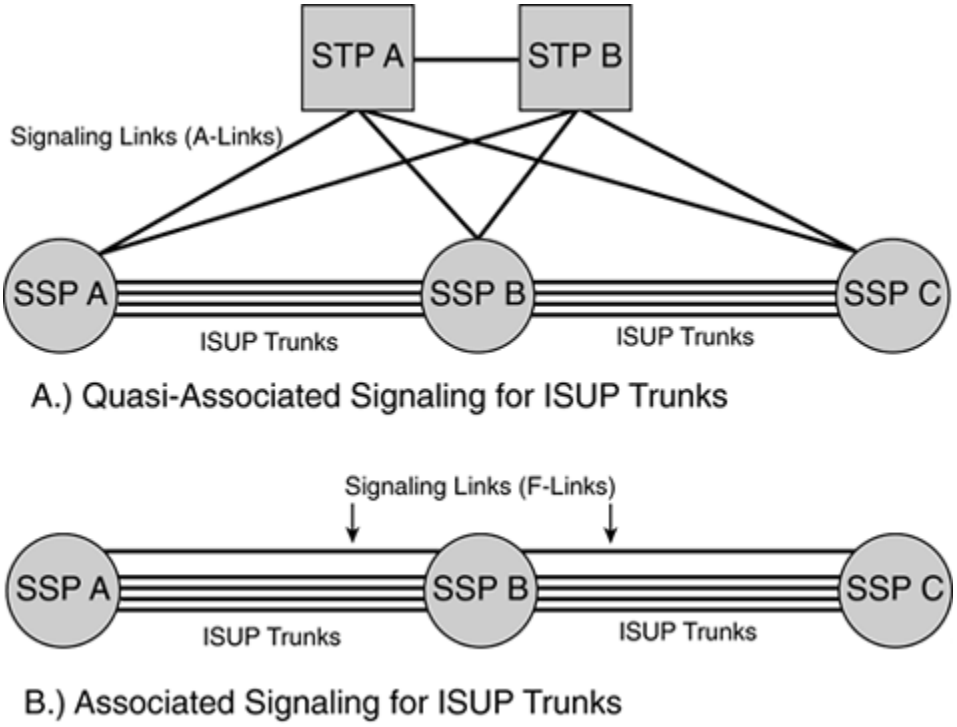
ISUP allows the call control signaling to be separated from the circuit that carries the voice stream over interoffice trunks. The circuit that carries the voice portion of the call is known within the telephone industry by many different terms. Voice channel, voice circuit, trunk member, and bearer all refer to the digital time slot that transports the voice (fax, modem, or other voiceband data) part of a call. The term "voice circuit" can be somewhat ambiguous in this context because sometimes it is used to refer to the trunk span that is divided into time slots, or to an individual time slot on a span.

The signaling component of the call is, of course, transported over SS7 signaling links. This creates two independent paths for call information between nodes: the voice path and the signaling path. The signaling mode describes the signaling relation between the two paths. Following is a brief review of the associated and quasi-associated signaling modes as they relate to ISUP, which we discussed in earlier chapters.

If the signaling travels on a single linkset that originates and terminates at the same nodes as the bearer circuit, the signaling mode is associated. If the signaling travels over two or

more linksets and at least one intermediate node, the signaling mode is quasi-associated. In Figure 8-1, part A shows quasi-associated signaling between SSP A and SSP B and between SSP B and SSP C. In part B of Figure 8-1, the same SSP nodes are shown using associated signaling. Notice that the signaling links in part B terminate at the same point as the trunks. Also, the signaling link is shown as a separate entity in part B to illustrate the signaling mode; however, it is typically just another time slot that is dedicated for signaling on a trunk span.

Figure 8-1. Signaling Mode Relating to ISUP Trunks

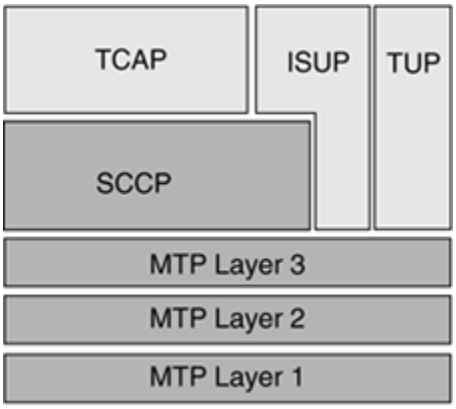


The signaling mode used for ISUP depends greatly on what SS7 network architecture is used. For example, North America uses hierarchical STPs for aggregation of signaling traffic. Therefore, most ISUP trunks are signaled using quasi-associated signaling. Using this mode, the signaling is routed through the STP before reaching the destination SSP. In contrast, while the U.K. uses quasi-associated signaling for some SSPs, they also heavily use associated signaling with directly connected signaling links between many SSPs.

ISUP and the SS7 Protocol Stack

As shown in Figure 8-2, ISUP resides at Level 4 of the SS7 stack with its predecessor, the Telephone User Part (TUP). TUP is still used in many countries, but ISUP is supplanting it over time. TUP also provides a call setup and release that is similar to ISUP, but it has only a subset of the capabilities. TUP is not used in North America because its capabilities are not sufficient to support the more complex network requirements.

Figure 8-2. ISUP at Level 4 of the SS7 Stack



As you can see in Figure 8-2, a connection exists between ISUP and both the SCCP and MTP3 levels. ISUP uses the MTP3 transport services to exchange network messages, such as those used for call setup and clear down. The connection to SCCP is for the transport of end-to-end signaling. While SCCP provides this capability, today ISUP end-to-end signaling is usually transported directly over MTP3. The "Interworking with ISDN" section of this chapter further discusses end-to-end signaling and the two different methods using MTP3 and SCCP for transport.

ISUP Standards and Variants

The ITU-T defines the international ISUP standards in the Q.767 and the national standards in the Q.761–Q.764 series of specifications. The ITU-T standards provide a basis from which countries or geographical regions can define regional or national versions of the protocol, which are often referred to as variants. For the U.S. network, the following standards provide the primary specifications for the ISUP protocol and its use in local and long distance networks:

- ANSI T1.113–ANSI ISUP
- Telcordia GR-246 Telcordia Technologies Specification of Signaling System No. 7, Volume 3. (ISUP)
- Telcordia GR-317 LSSGR— Switching System Generic Requirements for Call Control Using the Integrated Services Digital Network User Part (ISDNUP)
- Telcordia GR-394 LSSGR— Switching System Generic Requirements for Interexchange Carrier Interconnection (ICI) Using the Integrated Services Digital Network User Part (ISDNUP)

In Europe, the following ETSI standards provide the basis for the national ISUP variants:

- ETSI ETS 300-121 Integrated Services Digital Network (ISDN); Application of the ISDN User Part (ISUP) of CCITT Signaling System No. 7 for international ISDN interconnections

- ETSI ETS 300-156-x Integrated Services Digital Network (ISDN); Signaling System No. 7; ISDN User Part (ISUP) for the international interface

The ETS 300-121 is version 1, and the ETS 300-156-x (where x represents an individual document number) is a suite of specifications that covers ETSI ISUP versions 2–4.

A multitude of different country requirements have created many ISUP variants. A few of the several flavors are Swedish ISUP, U.K. ISUP, Japanese ISUP, Turkish ISUP, Korean ISUP. Each variant is tailored to the specific national requirements. Although not certain of the exact number of variants that are in existence today, the author has encountered over a hundred different ISUP variants while developing software for switching platforms.

ISUP Message Flow

This section provides an introduction to the core set of ISUP messages that are used to set up and release a call. The ISUP protocol defines a large set of procedures and messages, many of which are used for supplementary services and maintenance procedures. While the ITU Q.763 ISUP standard defines nearly fifty messages, a core set of five to six messages represent the majority of the ISUP traffic on most SS7 networks. The basic message flow that is presented here provides a foundation for the remainder of the chapter. Additional messages, message content, and the actions taken at an exchange during message processing build upon the foundation presented here.

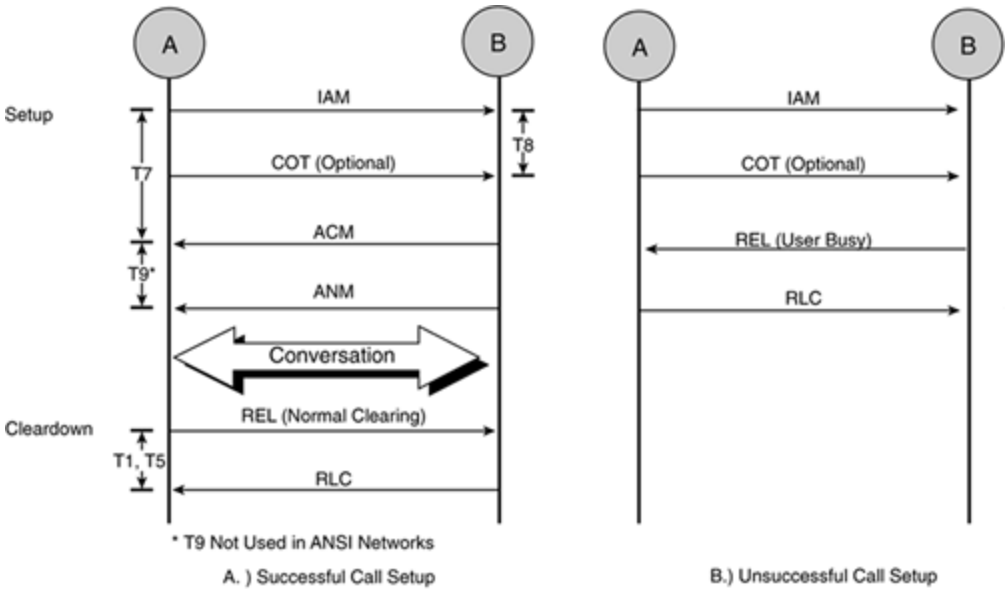
A basic call can be divided into three distinct phases:

- Setup
- Conversation (or data exchange for voice-band data calls)
- Release

ISUP is primarily involved in the set-up and release phases. Further ISUP signaling can take place if a supplementary service is invoked during the conversation phase.

In Figure 8-3, part A illustrates the ISUP message flow for a basic call. The call is considered basic because no supplementary services or protocol interworking are involved. The next section, "Call Setup," explains the figure's message timer values.

Figure 8-3. Simple ISUP Message Flow



Call Setup

A simple basic telephone service call can be established and released using only five ISUP messages. In Figure 8-3, part A shows a call between SSP A and SSP B. The Initial Address Message (IAM) is the first message sent, which indicates an attempt to set up a call for a particular circuit. The IAM contains information that is necessary to establish the call connection—such as the call type, called party number, and information about the bearer circuit. When SSP B receives the IAM, it responds with an Address Complete Message (ACM). The ACM indicates that the call to the selected destination can be completed. For example, if the destination is a subtending line, the line has been determined to be in service and not busy. The Continuity message (COT), shown in the figure, is an optional message that is used for continuity testing of the voice path before it is cut through to the end users. This chapter's "Continuity Test" section discusses the COT message.

Once the ACM has been sent, ringing is applied to the terminator and ring back is sent to the originator. When the terminating set goes off-hook, an Answer Message (ANM) is sent to the originator. The call is now active and in the talking state. For an ordinary call that does not involve special services, no additional ISUP messages are exchanged until one of the parties signals the end of the call by going on-hook.

Call Release

In Figure 8-3, the call originator at SSP A goes on-hook to end the call. SSP A sends a Release message (REL) to SSP B. The REL message signals the far end to release the bearer channel. SSP B responds with a Release Complete message (RLC) to acknowledge the REL message. The RLC indicates that the circuit has been released.

If the terminating party goes on-hook first, the call might be suspended instead of being released. Suspending a call maintains the bearer connection for a period of time, even though the terminator has disconnected. The terminator can go off-hook to resume the call,

providing that he does so before the expiration of the disconnect timer or a disconnect by the originating party. This chapter discusses suspending and resuming a connection in more detail in the section titled "Circuit Suspend and Resume."

NOTE

Several different terms are used to identify the two parties who are involved in a telephone conversation. For example, the originating party is also known as the calling party, or the "A" party. The terminating party, or "B" party, are also synonymous with the called party.

Unsuccessful Call Attempt

In Figure 8-3, part B shows an unsuccessful call attempt between SSP A and SSP B. After receiving the IAM, SSP B checks the status of the destination line and discovers that it is busy. Instead of an ACM, a REL message with a cause value of User Busy is sent to SSP A, indicating that the call cannot be set up. While this example shows a User Busy condition, there are many reasons that a call set-up attempt might be unsuccessful. For example, call screening at the terminating exchange might reject the call and therefore prevent it from being set up. Such a rejection would result in a REL with a cause code of Call Rejected.

NOTE

Call screening compares the called or calling party number against a defined list of numbers to determine whether a call can be set up to its destination.

Message Timers

Like other SS7 protocol levels, ISUP uses timers as a safeguard to ensure that anticipated events occur when they should. All of the timers are associated with ISUP messages and are generally set when a message is sent or received to ensure that the next intended action occurs. For example, when a REL message is sent, Timer T1 is set to ensure that a RLC is received within the T1 time period.

ITU Q.764 defines the ISUP timers and their value ranges. In Figure 8-3, part A includes the timers for the messages that are presented for a basic call. The "Continuity Test" section of this chapter discusses the timers associated with the optional COT message. Following are the definitions of each of the timers in the figure:

- **T7 awaiting address complete timer**— Also known as the network protection timer. T7 is started when an IAM is sent, and is canceled when an ACM is received. If T7 expires, the circuit is released.
- **T8 awaiting continuity timer**— Started when an IAM is received with the Continuity Indicator bit set. The timer is stopped when the Continuity Message is received. If T8 expires, a REL is sent to the originating node.

- **T9 awaiting answer timer**— Not used in ANSI networks. T9 is started when an ACM is received, and is canceled when an ANM is received. If T9 expires, the circuit is released. Although T9 is not specified for ANSI networks, answer timing is usually performed at the originating exchange to prevent circuits from being tied up for an excessive period of time when the destination does not answer.
- **T1 release complete timer**— T1 is started when a REL is sent and canceled when a RLC is received. If T1 expires, REL is retransmitted.
- **T5 initial release complete timer**— T5 is also started when a REL is sent, and is canceled when a RLC is received. T5 is a longer duration timer than T1 and is intended to provide a mechanism to recover a nonresponding circuit for which a release has been initiated. If T5 expires, a RSC is sent and REL is no longer sent for the nonresponding circuit. An indication of the problem is also given to the maintenance system.

We list the timers for the basic call in part A of Figure 8-3 to provide an understanding of how ISUP timers are used. There are several other ISUP timers; a complete list can be found in Appendix H, "ISUP Timers for ANSI/ETSI/ITU-T Applications."

Circuit Identification Codes

One of the effects of moving call signaling from CAS to Common Channel Signaling (CCS) is that the signaling and voice are now traveling on two separate paths through the network. Before the introduction of SS7 signaling, the signaling and voice component of a call were always transported on the same physical facility. In the case of robbed-bit signaling, they are even transported on the same digital time slot of that facility.

The separation of signaling and voice create the need for a means of associating the two entities. ISUP uses a Circuit Identification Code (CIC) to identify each voice circuit. For example, each of the 24 channels of a T1 span (or 30 channels of an E1 span) has a CIC associated with it. When ISUP messages are sent between nodes, they always include the CIC to which they pertain. Otherwise, the receiving end would have no way to determine the circuit to which the incoming message should be applied. Because the CIC identifies a bearer circuit between two nodes, the node at each end of the trunk must define the same CIC for the same physical voice channel.

TIP

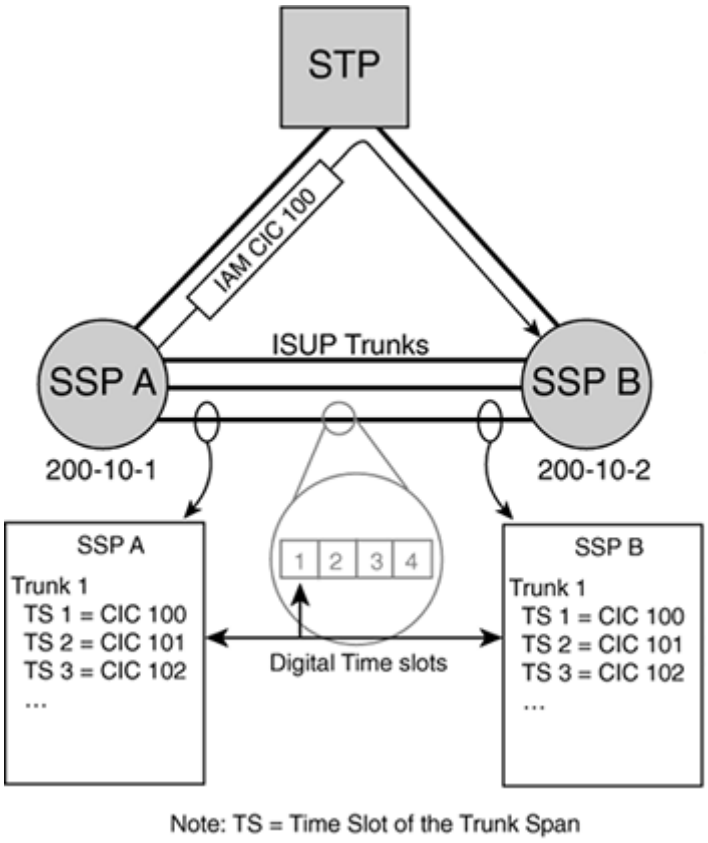
Not defining CICs so that they match properly at each end of the connection is a common cause of problems that occur when defining and bringing new ISUP trunks into service.

ITU defines a 12-bit CIC, allowing up to 4096 circuits to be defined. ANSI uses a larger CIC value of 14 bits, allowing for up to 16,384 circuits.

Figure 8-4 shows an ISUP message from SSP A that is routed through the STP to SSP B. For simplicity, only one STP is shown. In the message, CIC 100 identifies the physical circuit between SSP A and B to which the message applies. Administrative provisioning at each of the nodes associates each time slot of the digital trunk span with a CIC. As shown in the

figure, Trunk 1, time slot (TS) 1 is defined at each SSP as CIC 100. Trunk 1, time slot 2 is defined as CIC 101, and so on.

Figure 8-4. CIC Identifies the Specific Voice Circuit

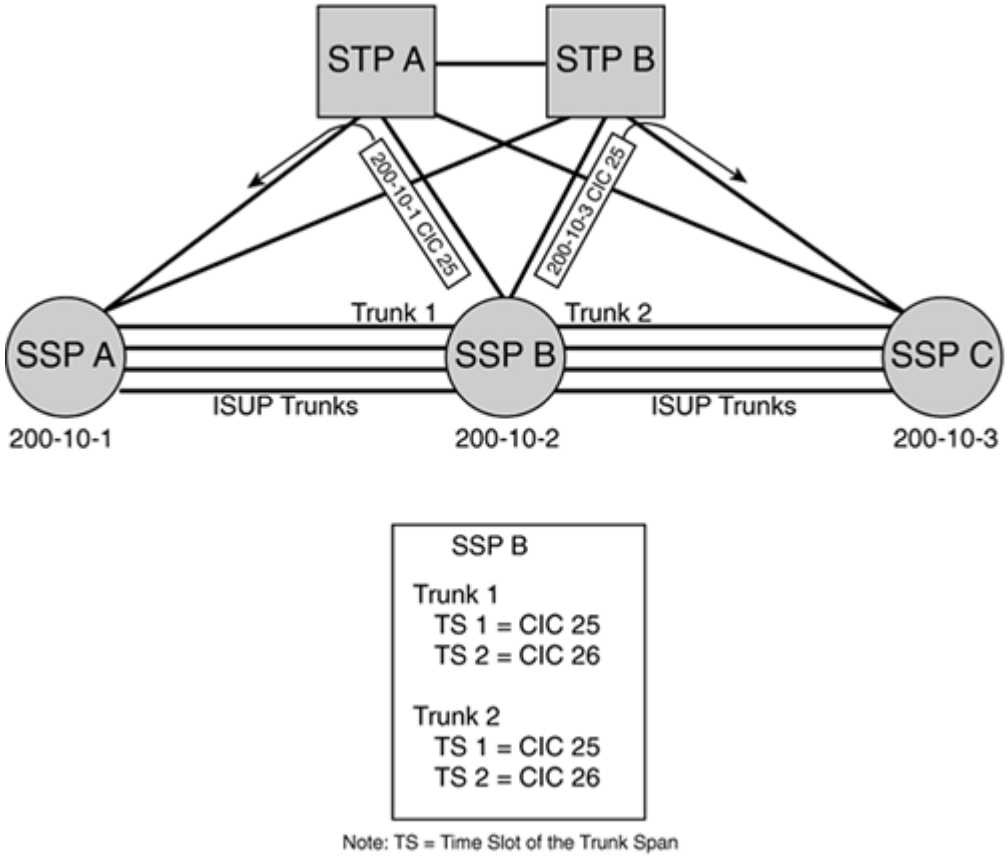


DPC to CIC Association

Since each ISUP message is ultimately transported by MTP, an association must be created between the circuit and the SS7 network destination. This association is created through provisioning at the SSP, by linking a trunk group to a routeset or DPC.

The CIC must be unique to each DPC that the SSP defines. A CIC can be used again within the same SSP, as long as it is not duplicated for the same DPC. This means that you might see CIC 0 used many times throughout an SS7 network, and even multiple times at the same SSP. It is the combination of DPC and CIC that uniquely identifies the circuit. Figure 8-5 shows an example of three SSPs that are interconnected by ISUP trunks. SSP B uses the same CIC numbers for identifying trunks to SSP A and SSP C. For example, notice that it has two trunks using CIC 25 and two trunks using CIC 26. Since SSP A and SSP C are separate destinations, each with their own unique routeset defined at SSP B, the DPC/CIC combination still uniquely identifies each circuit. SSP B can, in fact, have many other duplicate CIC numbers associated with different DPCs.

Figure 8-5. Combination of DPC/CIC Provide Unique Circuit ID



Unidentified Circuit Codes

When a message is received with a CIC that is not defined at the receiving node, an Unequipped Circuit Code (UCIC) message is sent in response. The UCIC message's CIC field contains the unidentified code. The UCIC message is used only in national networks

Enbloc and Overlap Address Signaling

The Called Party Number (CdPN) is the primary key for routing a call through the network. When using ISUP to set up a call, the CdPN can be sent using either enbloc or overlap signaling. In North America, enbloc signaling is always used; in Europe, overlap signaling is quite common, although both methods are used.

Enbloc Signaling

The enbloc signaling method transmits the number as a complete entity in a single message. When using enbloc signaling, the complete number is sent in the IAM to set up a

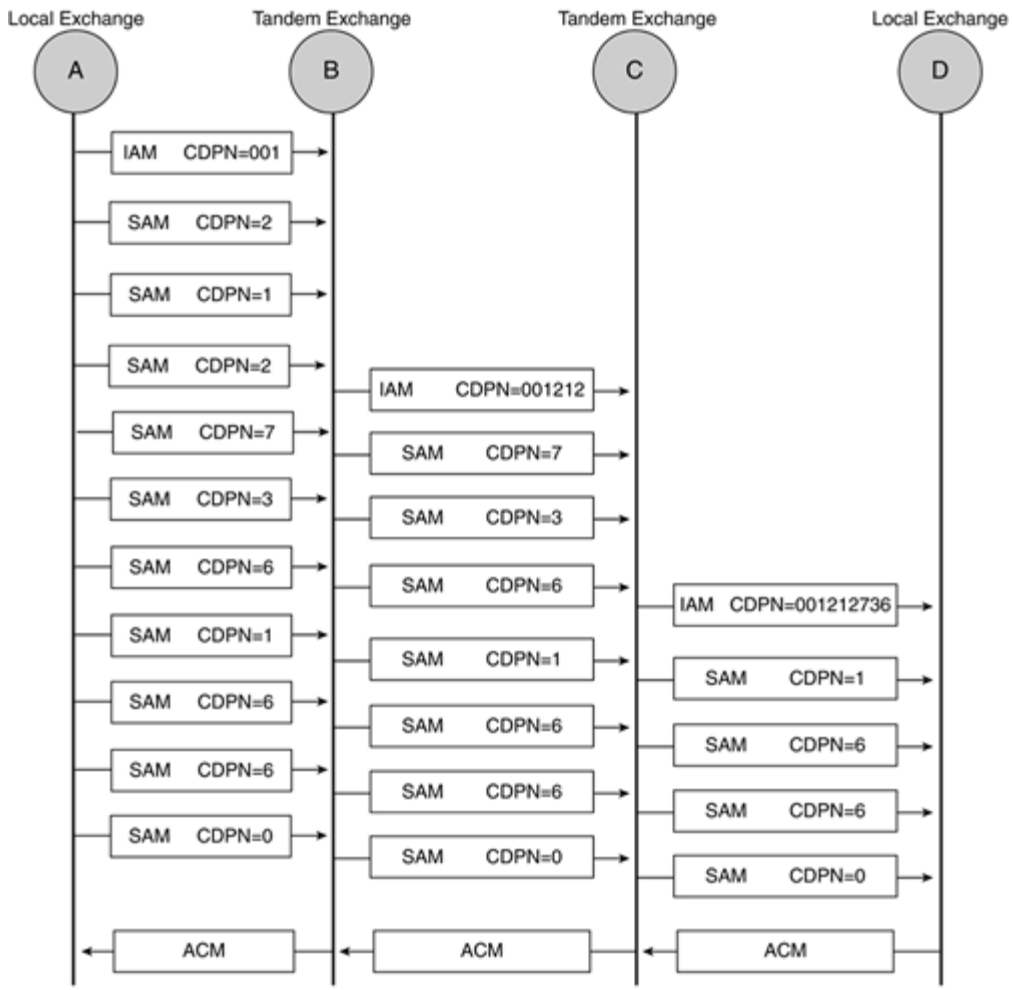
call. This is much more efficient than overlap signaling, which uses multiple messages to transport the number. Enbloc signaling is better suited for use where fixed-length dialing plans are used, such as in North America. Figure 8-6 illustrates the use of enbloc signaling.

Figure 8-6. Enbloc Address Signaling

Overlap Signaling

Overlap signaling sends portions of the number in separate messages as digits are collected from the originator. Using overlap signaling, call setup can begin before all the digits have been collected. When using the overlap method, the IAM contains the first set of digits. The Subsequent Address Message (SAM) is used to transport the remaining digits. Figure 8-7 illustrates the use of overlap signaling. Local exchange A collects digits from the user as they are dialed. When enough digits have been collected to identify the next exchange, an IAM is sent to exchange B. When tandem exchange B has collected enough digits to identify the next exchange, it sends an IAM to exchange C; exchange C repeats this process. After the IAM is sent from exchange C to exchange D, the destination exchange is fully resolved. Exchange D receives SAMs containing the remaining digits needed to identify the individual subscriber line.

Figure 8-7. Overlap Address Signaling



When using dialing plans that have variable length numbers, overlap signaling is preferable because it decreases post-dial delay. As shown in the preceding example, each succeeding call leg is set up as soon as enough digits have been collected to identify the next exchange.

As discussed in Chapter 5, "The Public Switched Telephone Network (PSTN)," interdigit timing is performed as digits are collected from a subscriber line. When an exchange uses variable length dial plans with enbloc signaling, it must allow interdigit timing to expire before attempting to set up the call. The exchange cannot start routing after a specific number of digits have been collected because that number is variable. By using overlap signaling, the call is set up as far as possible, waiting only for the final digits the subscriber dials. Although overlap signaling is less efficient in terms of signaling bandwidth, in this situation it is more efficient in terms of call set-up time.

Circuit Glare (Dual-Seizure)

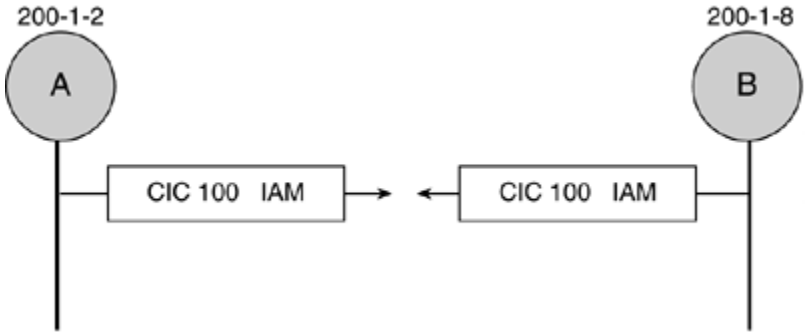
Circuit glare (also known as dual-seizure) occurs when the node at each end of a two-way trunk attempts to set up a call over the same bearer at the same time. Using ISUP signaling, this occurs when an IAM for the same CIC is simultaneously sent from each end. Each end sends an IAM to set up a call before it receives the IAM from the other end. You will recall from our discussion of the basic ISUP message flow that once an IAM is sent, an ACM is expected. When an IAM is received after sending an IAM for the same CIC, glare has occurred.

Resolving Glare

When glare is detected, one node must back down and give control to the other end. This allows one call to complete, while the other call must be reattempted on another CIC. There are different methods for resolving which end takes control. For normal 64-kb/s connections, two methods are commonly used. With the first method, the point code and CIC numbers are used to determine which end takes control of the circuit. The node with the higher-numbered point code takes control of even number CICs, and the node with the lower-numbered point code takes control of odd numbered CICs. This provides a fair mechanism that allows each node to control approximately half of the calls encountering glare. In the United States, an example of this use would be two peer End Office exchanges. The second method of glare resolution is handled by prior agreement between the two nodes about which end will back down when glare occurs. One node is provisioned to always back down, while the other node is provisioned to take control. A typical example of this arrangement in the U.S. network would be a hand-off between non-peer exchanges, such as an IXC to AT. The method to use for glare resolution can usually be provisioned at the SSP, typically at the granularity level of the trunk group.

Figure 8-8 illustrates a glare condition when SSP A and B have both sent an IAM before receiving the IAM from the other end. Assuming that the point code/CIC method of resolving glare is being used, SSP B takes control of the circuit because the CIC is even numbered and SSP B has a numerically higher point code.

Figure 8-8. Glare Condition During Call Setup



Avoiding Glare

When provisioning trunks, glare conditions can be minimized by properly coordinating the trunk selection algorithms at each end of a trunk group. A common method is to perform trunk selection in ascending order of the trunk member number at one end of the trunk group, and in descending order at the other end. This minimizes contention to the point of selecting the last available resource between the two ends. Another method is to have one end use the "Most Idle" trunk selection while the other end uses the "Least Idle" selection. The idea is to have an SSP select a trunk that is least likely to be selected by the SSP at the other end of the trunk group

Continuity Test

Continuity testing verifies the physical bearer facility between two SSPs. When CAS signaling is used, a call setup fails if the voice path is faulty. Using ISUP signaling, it is possible to set up a call using the signaling network without knowing that the bearer connection is impaired or completely broken.

The voice and signaling channels are usually on separate physical facilities, so a means of verifying that the voice facility is connected properly between the SSPs is needed. Many digital voice transmission systems provide fault detection on bearer facilities, which are signaled to the connected switching system using alarm indication bits within the digital information frame. However, these bits are not guaranteed to be signaled transparently through interconnecting transmission equipment, such as a Digital Access Cross Connect system (DACS) or digital multiplexers. Some networks require these alarm indications to be passed through without disruption, therefore, reducing the need for continuity testing.

Continuity testing can be considered part of the ISUP maintenance functions. It can be invoked to test trunks manually, as part of routine maintenance and troubleshooting procedures. Continuity testing can also be provisioned to take place during normal call setup and it has an impact on the flow of call processing. During call processing, the originating exchange determines whether a continuity test should be performed. Network guidelines vary concerning whether and how often continuity testing is performed. The determination is typically based on a percentage of call originations. For example, in the United States, the generally accepted practice is to perform continuity testing on 12 percent of ISUP call originations (approximately one out of eight calls). This percentage is based on Telcordia recommendations.

Loopback and Transceiver Methods

The actual circuit testing can be performed using either the loopback or the transceiver method. The loopback method is performed on four-wire circuits using a single tone, and the transceiver method is used for two-wire circuits using two different tones. The primary difference between the two methods is related to the action that takes place at the terminating end. When using either method, a tone generator is connected to the outgoing circuit at the originating exchange. Using the loopback method, the terminating exchange connects the transmit path to the receive path, forming a loopback to the originator. The originator measures the tone coming back to ensure that it is within the specified parameters. When the transceiver method is used, the transmit and receive path are connected to a tone transceiver that measures the tone coming from the originating

exchange and sends a different tone back to the originating exchange. The tone frequencies vary between countries. The following tones are used for the continuity test in North America:

- 2010 Hz from the originating exchange
- 1720 Hz from the terminating exchange (transceiver method only)

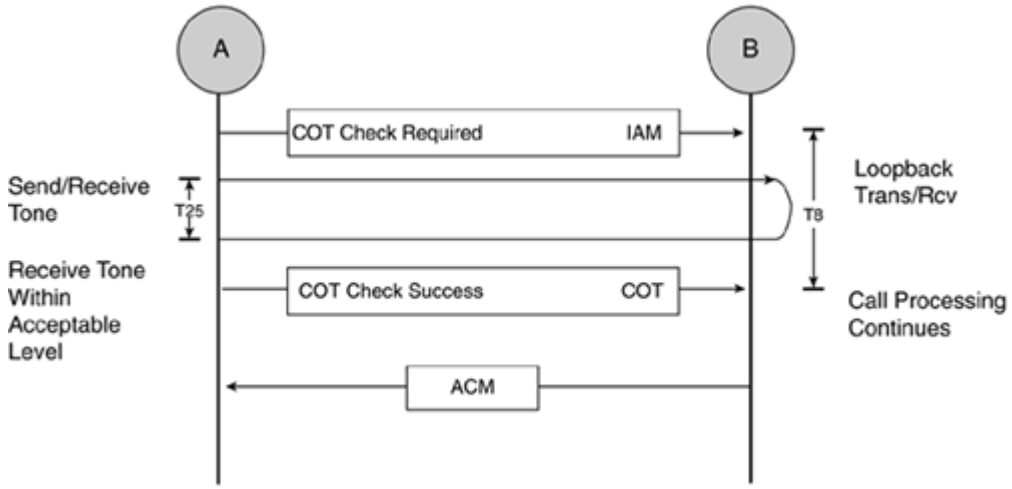
Another example of the COT tone frequency is 2000 Hz, which is used in the U.K.

Continuity Check Procedure

The Initial Address Message contains a *Continuity Check Indicator* as part of the *Nature of Connection* field. When an ISUP trunk circuit is selected for an outgoing call and the exchange determines that a continuity check should be performed, the Continuity Check Indicator is set to true. A tone generator is connected to the outgoing circuit, and the IAM is sent to the SSP at the far end of the trunk. Timer T25 is started when the tone is applied, to ensure that tone is received back within the T25 time period. When the SSP at the far end receives the IAM with the Continuity Check Indicator set to true, it determines whether to create a loopback of the transmit and receive path, or to connect a transceiver. The transceiver receives the incoming tone and generates another tone on the outgoing circuit. The determination of whether to use a loopback or transceiver is typically based on provisioned data at the receiving exchange. Upon receipt of the IAM, Timer T8 is started at the terminating exchange, awaiting the receipt of a COT message to indicate that the test passed. The terminating exchange does not apply ringing to the called party or send back ACM until the COT message has been received with a continuity indicator of continuity check successful to indicate that the bearer connection is good.

The originating exchange measures the received tone to ensure that it is within an acceptable frequency range and decibel level. Next it sends a COT message to the terminating exchange to indicate the test results. If the test passes, the call proceeds as normal; if the test fails, the CIC is blocked, the circuit connection is cleared, and the originating exchange sends a Continuity Check Request (CCR) message to request a retest of the failed circuit. While ISUP maintenance monitors the failed circuit's retest, ISUP call processing sets the call up on another circuit. Figure 8-9 shows a successful COT check using the loopback method.

Figure 8-9. Successful COT Check Using the Loopback Method



ISUP Message Format

The User Data portion of the MTP3 Signaling Information Field contains the ISUP message, identified by a Service Indicator of 5 in the MTP3 SIO field. Each ISUP message follows a standard format that includes the following information:

- **CIC**— the Circuit Identification Code for the circuit to which the message is related.
- **Message Type**— the ISUP Message Type for the message (for example, an IAM, ACM, and so on).
- **Mandatory Fixed Part**— required message parameters that are of fixed length.
- **Mandatory Variable Part**— required message parameters that are of variable length. Each variable parameter has the following form:

  - Length of Parameter

  - Parameter Contents

Because the parameter is not a fixed length, a field is included to specify the actual length.

- **Optional Part**— Optional fields that can be included in the message, but are not mandatory. Each optional parameter has the following form:

  - Parameter Name

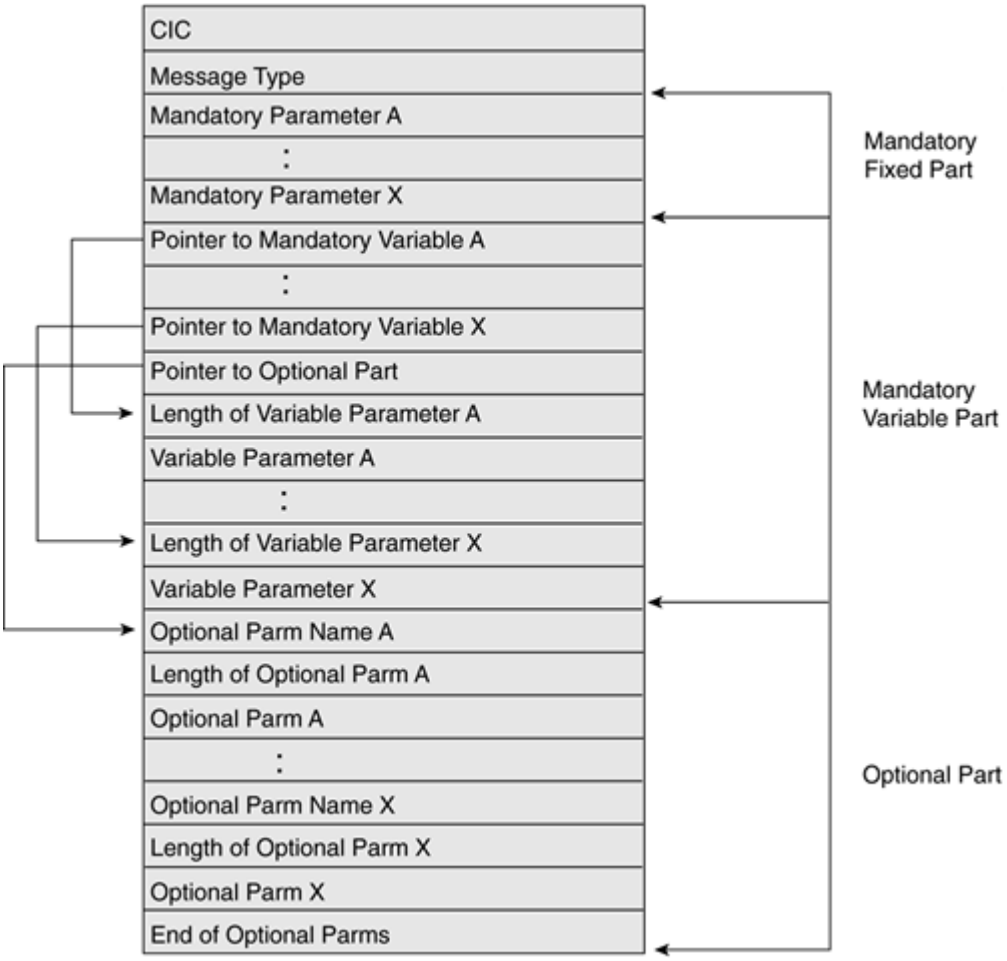
  - Length of Parameter

  - Parameter Contents

Figure 8-10 shows the ISUP message structure, as described here. This message structure provides a great deal of flexibility for constructing new messages. Each message type defines the mandatory parameters that are necessary for constructing a message. The mandatory fixed variables do not contain length information because the ISUP standards specify them to be a fixed length. Because the mandatory variable parameters are of variable lengths, pointers immediately follow the mandatory fixed part to point to the

beginning of each variable parameter. The pointer value is simply the number of octets from the pointer field to the variable parameter length field.

Figure 8-10. ISUP Message Format



In addition to the mandatory fields, each message can include optional fields. The last of the pointer fields is a pointer to the optional part. Optional fields allow information to be included or omitted as needed on a per-message basis. The optional fields differ based on variables such as the call type or the supplementary services involved. For example, the Calling Party Number (CgPN) field is an optional parameter of the IAM, but is usually included to provide such services as Caller ID and Call Screening.

A single message can include many optional parameters. The optional part pointer field only points to the first parameter. Because the message might or might not include the parameters, and because the parameters can appear in any order, the first octet includes the name of each parameter in order to identify it. The parameter length follows the name to indicate how many octets the parameter contents include. When the parameter name is coded as zero, it signals the end of the optional parameters. During parsing of an incoming ISUP message, optional parameters are processed until the *end of optional parameters* marker is reached. If the message does not have any optional parameters, the pointer to the optional part is coded to zero.

Basic Call Message Formats

Here, we examine the six messages shown in the basic call setup because they comprise the core message set for basic call setup and release, and are therefore used frequently. There are slight variations in the messages used based on the individual network. For example, Europe uses the SAM frequently and the COT message more rarely. In North America, SAM is not used at all, but COT is used more often. This section considers the following messages:

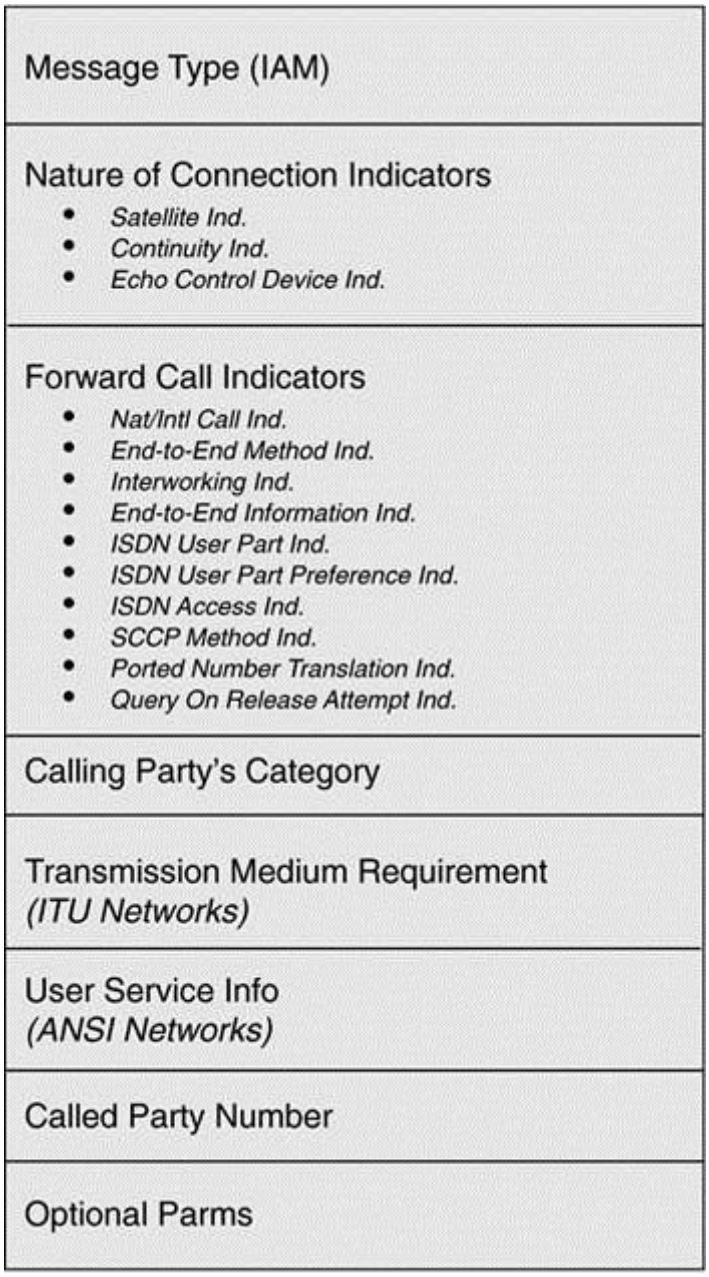
- Initial Address Message (IAM)
- Subsequent Address Message (SAM–ITU Networks only)
- Continuity Message (COT)
- Address Complete Message (ACM)
- Answer Message (ANM)
- Release Message (REL)
- Release Complete Message (RLC)

The following sections show only the mandatory fields for each message. Keep in mind that many optional parameters can also be included. In each of the figures, the fixed mandatory fields with sub-fields have been expanded to show what they are. For the sake of brevity in the figures, the variable subfields have not been expanded. All of the ISUP Message formats and parameters are documented in ITU-T Q.763. ANSI T1.113 documents the North American version of the messages.

Initial Address Message (IAM)

The IAM contains the information needed to set up a call. For a basic call, it is the first message sent and is typically the largest message in terms of size. Figure 8-11 shows the mandatory fields that the message includes. In addition to the mandatory fields, the ITU-T Q.764 lists more than 50 optional parameters that can be included in the IAM. The mandatory parameters for ITU and ANSI are the same, with the exception of the Transmission Medium Requirements parameter. In ANSI networks, the User Service Info field is used instead.

Figure 8-11. IAM Message Format



As shown in Figure 8-11, the *Nature of Connection Indicators (NOC)* pass information about the bearer circuit connection to the receiving node. The indicators consist of the following subfields:

- **Satellite Indicator**— Specifies whether one or more satellites have been used for the circuit connection that is being set up. This information is useful when setting up calls to prevent an excessive number of satellite hops, which can reduce the quality of calls.

- **Continuity Indicator**— Designates whether to perform a continuity check on the circuit being set up.
- **Echo Control Device Indicator**— Specifies whether echo suppression is used on the circuit. Echo suppression is used to increase the quality of voice calls by reducing echo, but it can damage data and fax calls because it subtracts a portion of the voice-band signal.

The *Forward Call Indicators (FCI)* contain information that specifies both the preferences about call setup in the forward direction and the conditions encountered so far in setting up the call. They include the following subfields:

- **National/International Call Indicator**— Indicates whether the call is coming in as National or International. International calls are specified by ITU international procedures, and national calls are processed according to national ISUP variant standards.
- **End-to-End Method Indicator**— Indicates the method used for signaling end-to-end information. SCCP and pass-along are the two end-to-end methods that are used. The pass-along method traverses each node in the connection to deliver information to the correct node. The SCCP method uses connectionless signaling to send information directly to its destination.
- **Interworking Indicator**— Indicates whether the connection has encountered interworking with non-SS7 facilities (for example, MF trunks). Interworking with non-SS7 facilities can limit or prohibit the capability of supplementary services or certain call types that require SS7 signaling.
- **End-to-End Information Indicator**— Indicates whether any end-to-end information is available.
- **ISDN User Part Indicator**— Indicates whether ISUP has been used for every leg of the connection. Note that this is not the same as the Interworking Indicator. It is possible to have an SS7-signaled circuit, but not use ISUP (for example, TUP signaling); however, if interworking has been encountered, this indicator is set to *ISDN User Part not used all the way*.
- **ISDN User Part Preference Indicator**— Specifies whether an ISUP facility is required or preferred when choosing an outgoing circuit. Some supplementary services or call types are not possible over non-ISUP facilities. If ISUP is required but not available, the call is released because the requested facility's preference cannot be met. If the preference indicator is set to *preferred*, an ISUP facility is chosen, if available; however, the call is still set up as long as a facility is available, even if it is not ISUP.
- **ISDN Access Indicator**— Indicates whether the originating access is ISDN or non-ISDN. ISDN provides a much richer interface to services that is not available on plain analog lines. This indicator suggests that the ISDN interface is available so that end-to-end signaling, backward requests for information, and so on can be carried out.
- **SCCP Method Indicator**— Indicates which method, if any, is used for SCCP end-to-end signaling. SCCP might use connection-oriented, connectionless, or both.

The *Calling Party's Category* specifies a general category into which the calling party is classified—such as an ordinary calling subscriber, operator, payphone, or test call.

The *Transmission Medium Requirement* (*TMR*) is not applicable to ANSI networks and is only supported in ITU-T networks. It contains the requirements for the bearer circuit capabilities (speech, 3.1-kHz audio, 64-Kb unrestricted, and so forth) that are needed for the call being

set up. For example, a video conference might require a 384-Kbs unrestricted circuit to guarantee an acceptable level of video quality.

User Service Information (*USI*) is used in ANSI networks instead of the ITU-T specified TMR. It contains the requirements for the bearer circuit capabilities (speech, 3.1-kHz audio, and 64-Kbs unrestricted) along with additional information such as layer 1 codec, circuit, or packet transfer mode and other bearer-related specifics.

The *Called Party Number (CdPN)* is the destination number that the calling party dials. The CdPN contains the following fields:

- **Odd/Even Indicator**— Indicates an odd or even number of digits in the CdPN.
- **Nature of Address Indicator**— Indicates the type of number (for example, National Significant Number or International). The receiving switch uses this indicator during translations to apply the number's proper dial plan.

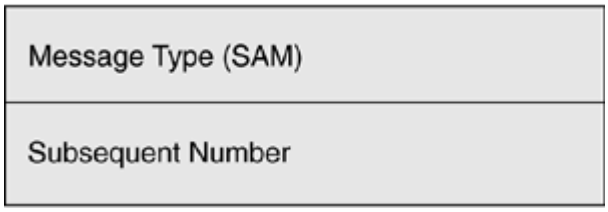
The *Internal Network Number Indicator* (*INN*), which is not used for ANSI, specifies whether routing to an internal network number is permitted. This field is used to block routing to specific numbers that should not be directly accessible from outside of the network. For example, if a premium rate number is translated to an internal number, the subscriber is blocked from dialing the internal number to ensure that the appropriate premium rate charges are collected.

- **Numbering Plan Indicator**— Specifies the type of number plan used. The E.164 ISDN numbering plan is commonly used for voice calls.
- **Address Signals**— The actual digits that comprise the called number. This includes digits 0–9 and the overdecadic digits (A–F), however, the overdecadic digits are not supported in all networks. Each digit is coded as a four-bit field.

Subsequent Address Message (SAM–ITU Networks Only)

Shown in Figure 8-12, the SAM is used to send subsequent address signals (digits) when using overlap signaling for call setup. It has one mandatory variable parameter: the *subsequent number*. One or more SAMs can be sent after an IAM to carry subsequent digits for call setup that are part of a destination's complete telephony number.

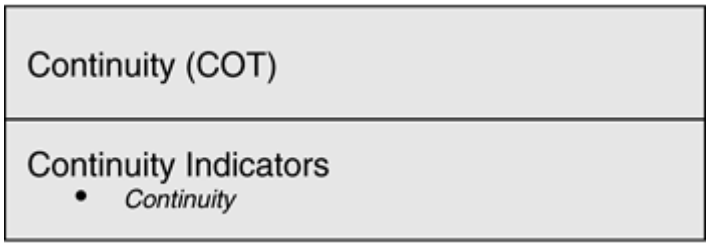
Figure 8-12. SAM Message Format



Continuity Message (COT)

As shown in Figure 8-13, the COT message contains the results of a continuity test. It has only one field: the *Continuity Indicators*. This field uses a single bit to indicate whether a continuity test passed or failed. The test's originator sends the message to the far end of the circuit that is being tested.

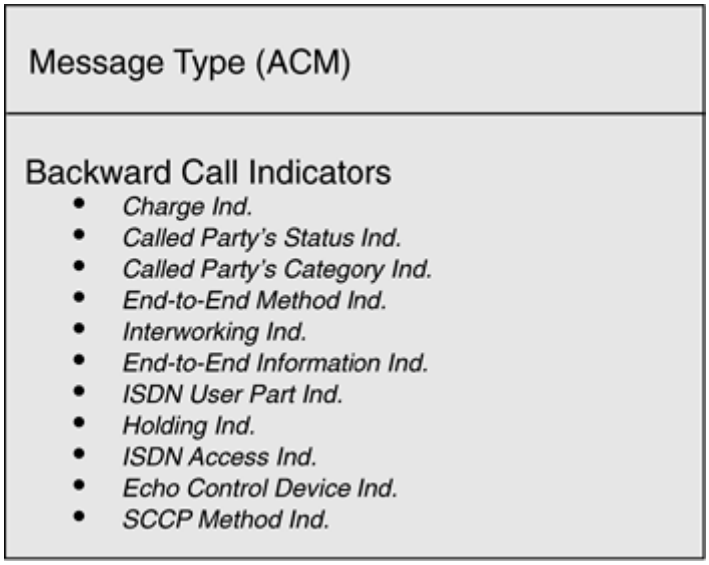
Figure 8-13. COT Message Format



Address Complete Message (ACM)

As shown in Figure 8-14, a destination node sends the ACM to indicate that a complete CdPN has been received. When enbloc signaling is used to set up the call, the ACM is sent after receiving the IAM; when overlap signaling is used, it is sent after the last SAM is received. In addition to indicating the successful reception of the CdPN, the ACM sends Backward Call Indicators (BCI) to signal information about the call setup. It is not mandatory for an ACM to be sent when setting up a call. It is permissible to send an ANM after receiving an IAM; this is sometimes referred to as "fast answer."

Figure 8-14. ACM Message Format



Many of the fields contained in the Backward Call Indicators are the same as those in the Forward Call Indicators (FCI), which are contained in the IAM. While the FCI signals the call indicators in the forward direction to provide information on the call setup to the terminating access (and intermediate nodes), the BCI signals similar information in the backward direction to the originator.

Here we discuss only the fields that are unique to the BCI. The remaining fields are the same as those we discussed for the FCI, except that they are representative of the call from the terminating end. For example, the ISDN Access Indicator specifies whether the "terminator" is ISDN.

- **Charge Indicator**— Indicates whether a call should be charged as determined by the charging exchange.
- **Called Party's Status Indicator**— Indicates whether the subscriber is free.
- **Called Party's Category Indicator**— Indicates the general category of the called party, an ordinary subscriber, or payphone.
- **Holding Indicator**— Indicates whether holding is required. This indicator can be used for special services, such as Operator Signaling Services or Malicious Call Tracing, to indicate that the incoming connection should be held. No specification for ANSI networks exists.

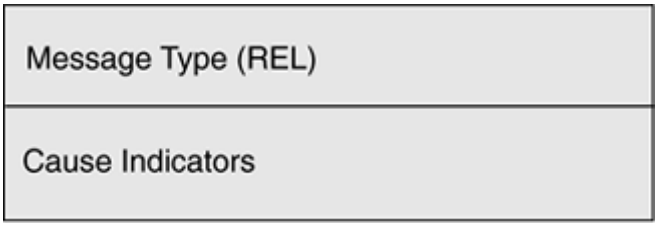
Answer Message (ANM)

The ANM is sent to the previous exchange when the called party answers (off-hook). Although it might contain many optional parameters, the ANM does not contain any mandatory fields other than the message type.

Release Message (REL)

As shown in Figure 8-15, the REL message indicates that the circuit is being released. When a RLC has been received in response, the circuit can be returned to the idle state for reuse. The REL message can be sent in either direction. It contains a single mandatory *Cause Indicators* field to indicate why the circuit is being released.

Figure 8-15. REL Message Format



Cause Indicators specify the cause information associated with the circuit being released. The Cause Indicators contain the general location in the network (such as local, remote, or transit) in which the circuit was released. The Coding Standard indicates which standard is used for decoding the Cause Value (such as ANSI, ITU). ANSI and ITU define some cause values differently, and ANSI also has additional values the ITU does not include.

The *Cause Value* contains an integer that represents the reason the circuit is being released. This value can be further decomposed into a class and a value. The most significant three bits of the Cause Value field represent the class. Each class is a general category of causes; for example, binary values of 000 and 001 are *normal event* class, and a value of 010 is *resource unavailable*. So, a cause value of 1 (unallocated number) is in the *normal event* class and a cause value of 34 (no circuit available) is in the*resource unavailable* class. Appendix M, "Cause Values," contains a complete list of the ITU and ANSI cause values.

The *Diagnostics* field is only applicable to certain cause values. It provides further information pertaining to the circuit release (for example, Transit Network Identity, Called Party Number [CdPN]) for those cause values.

Release Complete Message (RLC)

The RLC message is sent to acknowledge a REL message. Upon receipt of an RLC, a circuit can return to the idle state.

Detailed Call Walk-Through

Earlier in this chapter, we presented an ISUP message flow in order to illustrate the exchange of messages to establish and release an ISUP call. Now that we have discussed more of the ISUP details, we will build on that illustration. This section provides more detail about the call processing that was driven by the ISUP message events used in the earlier example. Although this chapter's primary focus is the ISUP protocol, it is important to understand how ISUP is applied in its normal domain of trunk call processing.

Call Setup

Refer back to Figure 8-3, where a call originates from a line at SSP A and terminates to a line at SSP B over an interexchange ISUP trunk. When call processing has completed translations of the called number at SSP A, the translations' results indicates that the call requires routing to an interexchange trunk group. The provisioned signaling type for the selected trunk group determines whether ISUP signaling or some other signaling, such as Multifrequency (MF), is used. When the signaling type is determined to be ISUP, the trunk circuit to be used for the outgoing call is reserved for use.

The SSP populates the IAM with information about the call setup, such as the CIC, CdPN, Call Type, CgPN, and PCM Encoding scheme. The IAM information is placed in the User Data field of the MTP3 SIF. The MTP3 information is populated based on the SS7 network information that is associated with the selected trunk group. As previously noted, each switching exchange contains a provisioned association (usually static) between routesets and trunkgroups. The IAM is then transmitted onto a signaling link toward the destination identified in the message by the DPC. If quasi-associated signaling is used, the message's

next-hop node is an STP that will route the message to the intended SSP. If associated signaling is used, the IAM is transmitted directly to the SSP that is associated with the trunk being set up. SSP A starts timer T7, which is known as the *network protection timer,* or the *awaiting ACM timer*, to ensure that an ACM is received in response to the IAM.

When SSP B receives the MTP3 message, it recognizes it as an ISUP message by the SIO's Service Indicator bit. Then the message is passed to ISUP for processing, during which it extracts the message information. An IAM indicates a request to set up a call so SSP B enters the call processing phase for a trunk origination. The CdPN and Calling Party Category fields provide key pieces of information from the IAM for SSP B to complete number translations for this simple call.

NOTE

The CdPN is commonly used to enter number translations processing; however, depending on call specifics, other fields can be used for translation. For example, calls involving ported numbers can use the Generic Address Parameter during number translation to determine the outgoing call destination.

In this example, the number translates to a subtending line of SSP B, which checks the line to determine whether it is available. An ACM is built and sent to SSP A, notifying that the call can be completed and is proceeding. At this point, the speech path in the backward direction (from SSP B to SSP A) should be cut through to allow the ring-back tone to be sent over the bearer channel from the terminating exchange to the originating exchange. This indicates that the terminator is being alerted.

NOTE

Note that the terminating office does not always send the ring-back tone. For example, ISDN can use the ACM message to notify the originating phone terminal to provide the ring-back tone.

Ringing is now applied to the terminating set, while ring back occurs at the originating set. Answer timing is usually applied at the originating switch to limit the amount of time an originator waits for answer.

When the terminating subscriber goes off-hook, an ANM is sent back to the originator to indicate that an answer has occurred. By this point, the voice path should be cut through in the forward direction to allow the conversation to take place. Note that the voice path can be cut through before receiving the ANM, but it must be cut through no later than the ANM. The call is now in the active, or talking, state. This is often a point of interest for billing procedures that require capturing the time at which a call conversation begins. For an ordinary call, no further signaling messages are exchanged for the duration of the conversation. When either of the parties goes on-hook, it initiates signaling for the release of the call. The following section discusses Call release.

Call Release

When either the originating or terminating subscriber goes on-hook, it signals an attempt to disconnect the call. In Figure 8-3, the originator at SSP A goes on-hook. SSP A recognizes the signal to disconnect the call and sends a Release message (REL) to SSP B. SSP B responds by sending a Release Complete message (RLC) as an acknowledgement. The trunk member is freed and placed back into its idle queue to be used for another call.

Terminal Portability

The ITU defines terminal portability in Q.733.4 for allowing the called or calling party to hang up a phone and resume a conversation at another phone that is connected to the same line. When the two parties are connected over an inter-exchange ISUP trunk, suspend and resume messages are used to maintain the trunk connection until the on-hook party has gone off-hook. Terminal portability requirements for the called party exist in many countries; however, terminal portability for the calling party is not supported as often. ANSI networks do not support terminal portability for the calling party.

Circuit Suspend and Resume

In Figure 8-3, the originating subscriber goes on-hook first. The originator is normally considered in control of the call, so the circuit is released when the originator goes on-hook. If the terminator goes on-hook while the originator remains off-hook, there are two methods of handling the disconnection.

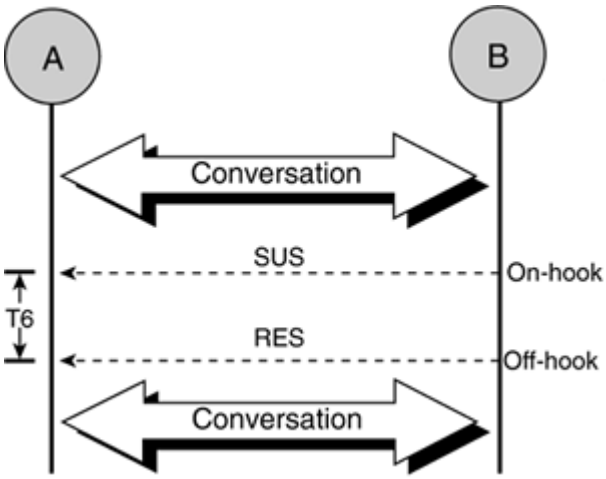
The first method is for the terminating exchange to release the call by sending a REL message to the originating exchange. This is no different than the scenario presented for a release initiated at the originating exchange; the originating switch responds with an RLC and the circuit is idled at each SSP.

The other method is for the terminating exchange to send a Suspend (SUS) message in the backward direction when it receives a disconnect indication from the terminating line. The SUS message provides notification that the terminating party has disconnected but that the circuit connection is still being maintained. Suspending the call allows the person who receives the call an opportunity to pick up on another phone extension.

When the SUS is received, the originating exchange starts a suspend timer (Timer T6, or Timer T38 in the case of an international exchange). If the terminating party reconnects (off-hook) before the suspend timer expires, a Resume (RES) message is sent in the backward direction, allowing the conversation to continue.

Figure 8-16 shows an example of a Suspend (SUS) and Resume (RES) being sent from the terminating exchange. If the suspend timer expires, a REL is sent in the forward direction. In the event that the originator goes on-hook during the time the circuit is suspended, the originating exchange sends a REL forward and normal call clearing takes place. The terminating exchange responds with a RLC.

Figure 8-16. ISUP Suspend/Resume



Support for SUS/RES varies, based on factors such as the type of service and the local network policies. For example, in the United States, SUS/RES is only supported for non-ISDN service

ISUP and Local Number Portability

Local Number Portability (LNP) is the concept of having phone numbers that remain the same for the subscriber, regardless of whether the subscriber changes service providers or geographic location. Historically, phone numbers have been associated with a particular geographic region or a particular service provider. The actual use of LNP in the network exists today, but only to a small degree. It is being expanded in phases and will take some time before it is ubiquitous across all networks and locations. This section examines the different mechanisms used to provide portability services and how these mechanisms relate to setting up calls with ISUP.

Chapter 11, "Intelligent Networks (IN)" provides an overview of the various phases identified under the umbrella of Number Portability (NP), such as service provider portability and location portability. Some of the mechanisms used for NP employ Intelligent Network (IN) databases, so we cover NP in part both in the Chapter 11 and in this chapter.

When NP is implemented, numbers are transitioned from physical addresses that identify an exchange location to virtual addresses that identify a subscriber. A means of mapping must be used to derive a physical address in the network from the called number because the number no longer identifies a physical destination. The network in which the physical number existed before portability was introduced is called a *donor* network. Each time a number is ported and becomes a virtual address, the network has "donated" a number that previously belonged to that network. We use the term "donor" or "donor network" several times during the discussion of NP. The network in which the physical number now resides is called the *recipient* network.

Currently, four mechanisms are defined for implementing NP:

- All Call Query (ACQ)
- Query on Release (QOR)
- Dropback or Release to Pivot (RTP)
- Onward Routing (OR)

Each method has its merits in terms of resource efficiencies, maintainability, and competitive fairness among network operators, but those topics are outside of the scope of the book. The details of how each mechanism is implemented also vary from country to country. The following section provides a general understanding of NP and how it affects the ISUP call flow and messages.

All Call Query (ACQ)

ACQ sends an IN query to a centrally administered database to determine the call's physical address or routing address. Chapter 11 discusses the details of the IN query. The way the routing number returned by the query is used varies based on national standards. The following example illustrates how the routing number is used in North America.

The number returned from the database is a Location Routing Number (LRN) that identifies the exchange serving the called number. Each exchange in the network is assigned an LRN. The IAM sent after the database query is performed contains the LRN in the CdPN field. The call is routed on the CdPN using switching translations to reach the destination exchange. The IAM also includes a Generic Address Parameter (GAP) with the original dialed number (the virtual address). This allows the destination exchange to set up the call to the intended subscriber because the LRN can only identify the exchange. The Forward Call Indicators of the IAM include a Ported Number Translation Indicator (PNTI), which indicates that a query for the ported number has been performed.

Query On Release (QOR)

QOR routes the call from the originator to the donor network's ported number in the same manner used prior to NP. The donor network releases the call back with a cause value of *Number Portability QOR number not found* (ITU causes value 14, ANSI causes value 27 in the REL message). The originating network then performs a query to an NP database to determine what routing number to use in the IAM in order to reach the recipient network.

Dropback (Also Known as Release to Pivot)

Dropback, or *Release to Pivot* (*RTP*), routes the call to the ported number in the donor network, just like QOR. However, instead of having the originating network query for the number, the donor exchange provides the routing number for the ported number when it releases back to the originator.

Onward Routing (OR)

Onward Routing (*OR*) also routes the call to the donor network's ported number. It differs from QOR and RTP in that it does not release the call back to the originating network. Rather, it references an internal database to determine the new routing number that is associated with the ported number and uses the new number to route the call.

Using the QOR and RTP mechanisms, an IAM is sent and an REL received back from the donor network, therefore, requiring a subsequent call attempt. The ACQ and OR do not release back or require subsequent call attempts. The OR mechanism creates additional call legs because the call is being connected through the donor network rather than being directly set up to the recipient network.

ISUP-ISUP Tandem Calls

Previous scenarios have focused on line-ISUP and ISUP-line calls. ISUP processing at a tandem switch occurs in the same sequence as the line to ISUP calls we discussed previously. However, in the case of ISUP-ISUP calls, the trigger for call processing events on the originating and terminating side are incoming ISUP messages.

This section discusses the following three areas that are related to ISUP processing at a tandem node:

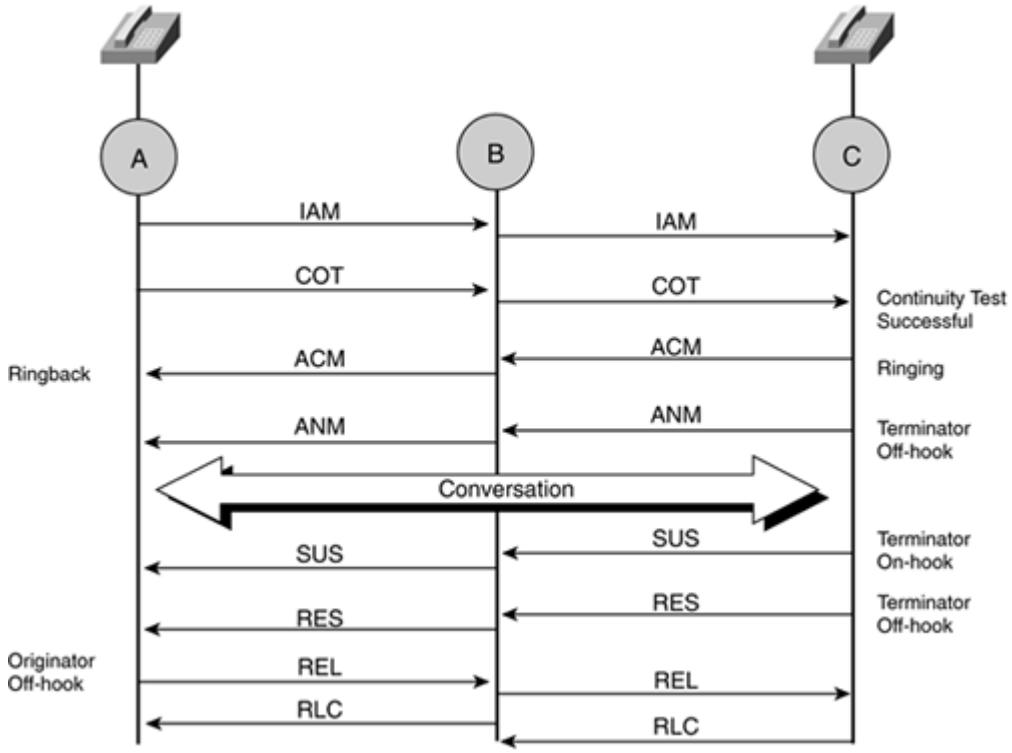
- ISUP Message Processing
- Continuity Testing
- Transporting Parameters

ISUP Message Processing at a Tandem

In Figure 8-17, the call origination at SSP B is based on an incoming ISUP origination (IAM) from another exchange. The fields that are necessary for number translation, such as CdPN, are extracted from the IAM and used to process the call at the tandem node to determine the outgoing destination. The translation and routing process results in the selection of an outgoing ISUP trunk. An IAM is sent in the forward direction to SSP C, updating fields in the message as necessary. For example, a new CdPN might be inserted as a result of translations. The NOC field is updated based on information such as whether a satellite is being used for the voice circuit or whether a continuity check is being performed.

Figure 8-17. ISUP-ISUP Tandem Calls

[View full size image]



When the ACM and ANM are received at SSP B, they are propagated to SSP A, updating fields such as the BCI as necessary. Each leg of the call cuts through the speech path in the same manner discussed in the "Detailed Call Walk-Through" section of this chapter.

When SSP A sends an REL message, SSP B responds with an RLC. It does not need to wait for the RLC to be sent from SSP C. Next, SSP B sends an REL to SSP C and waits for RLC to complete the release of that leg of the call. Keep in mind that even though some messages in a multi-hop ISUP call are propagated, the entire call actually consists of independent circuit segments. The release procedure is a reminder of this fact because the RLC can be sent immediately after receiving a REL.

Continuity Testing

When a call is set up across multiple exchanges, continuity testing is performed independently on each leg of the call. If a call traverses three trunks across four different exchanges and continuity is done on a statistical basis, it will likely only be performed on some of the trunks involved in the call. While the actual continuity test is performed independently on each call leg, the end-to-end call setup is dependent on each leg passing the test. If a continuity test is successfully performed on the second leg of the call (SSP B to SSP C), the results are not reported until the COT results have been received from the previous leg of the call (SSP A to SSP B). If a previous leg of the call connection cannot be set up successfully, there is no need to continue. For example, if SSP A reports a COT failure, it would attempt to establish a new connection in the forward direction by selecting another circuit to set up the call. There is no need to continue the previous connection from

SSP B to SSP C because the new call attempt from SSP A will come in as a new origination to SSP B.

Transporting Parameters

A tandem node can receive ISUP parameters that are only of interest to the destination exchange. This is particularly true of many optional parameters, which are passed transparently in the outgoing messages across tandem nodes. However, the tandem might update some fields during call processing, based on new information encountered while processing. For example, a tandem node that selects an outgoing ISUP facility over a satellite connection would update the NOC *Satellite Indicator* field in the outgoing IAM. This distinction is made because the tandem node might be required to have knowledge of how to process some parameters, but not others. When parameters are passed across a tandem node without processing the information, it is sometimes referred to as "ISUP transparency." Since the parameters do not need to be interpreted by the tandem, they are considered transparent and are simply relayed between the two trunks

Interworking with ISDN

ISDN uses a common channel (the D channel) for access signaling; this compliments the common channel network signaling ISUP uses and provides a complete digital signaling path between end users when ISDN is used for network access and ISUP is used throughout the core network. The ISUP/ISDN interworking specifications for ITU-T, ETSI, and Telcordia are found in the following standards:

- ITU-T Q.699—Interworking of Signaling Systems—Interworking Between Digital Subscriber Signaling System No. 1 and Signaling System No. 7
- ETSI EN 300-899-1 Integrated Services Digital Network (ISDN); Signaling System No. 7; Interworking Between ISDN User Part (ISUP) Version 2 and Digital Subscriber Signaling System No. one (DSS1); Part 1: Protocol Specification
- Telcordia GR-444 Switching System Generic Requirements Supporting ISDN Access Using the ISDN User Part

A correlation exists between the ISDN messages from the user premises and the ISUP messages on the network side of the call. Figure 8-18 illustrates this correlation using an ISDN-to-ISDN call over an ISUP facility. Table 8-1 lists the message mapping that occurs between the two protocols for the basic call setup shown in the diagram.

Figure 8-18. ISUP-ISDN Interworking

[View full size image]

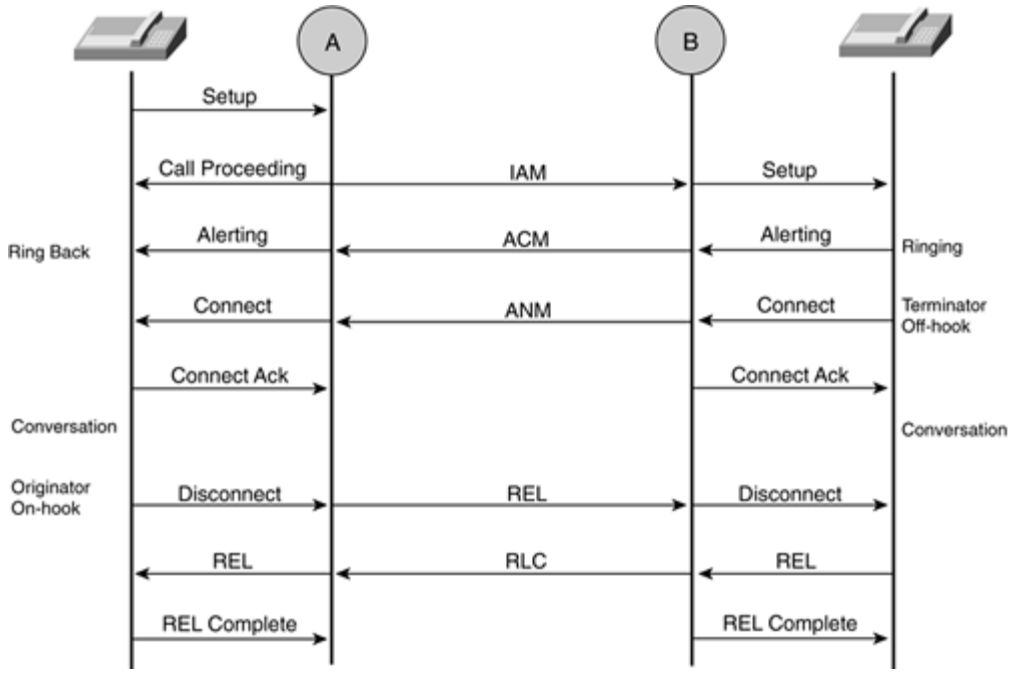


Table 8-1. Message Mapping Between ISDN and ISUP

ISDN	ISUP
Setup	IAM
Alerting	ACM (or CPG)
Connect	ANM (or CON)
Disconnect	REL
Release	RLC

Many of the fields within these messages also have direct mappings. For example, the bearer capability field in the ISDN Setup message maps to the ANSI User Service Info or the IAM's ITU Transmission Medium Requirements field. There are fields that have no direct mapping, such as the NOC Indicators and FCIs in the IAM. Many of the fields that do not have direct mapping contain network-specific information that would not be useful for the ISDN signaling endpoint.

End-to-End Signaling

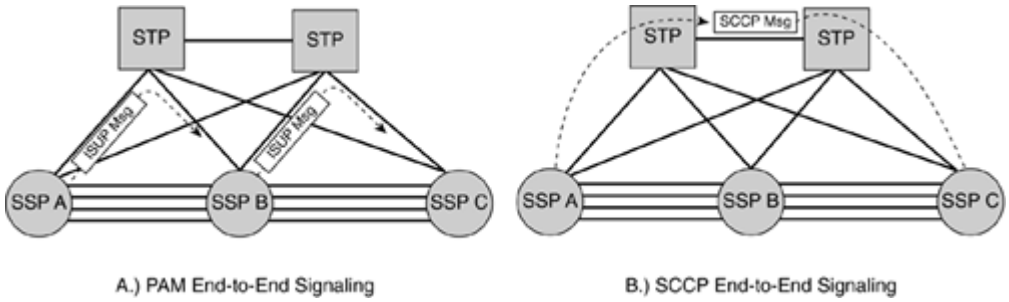
The ability to perform end-to-end signaling is accomplished using ISDN access signaling and ISUP network signaling. End-to-end signaling is the passing of information across the

network that is only pertinent to the two communicating endpoints. Generally, this means that the two phone users are connected across the network. The network itself can be viewed as a communications pipe for the user information.

There are two different methods for end-to-end signaling over ISUP: the *Pass Along Method (PAM)* and the *SCCP Method*. As shown in Figure 8-19, PAM exchanges end-to-end signaling by passing along information from one node to the next, based on the physical connection segments. The SCCP method uses a call reference to pass end-to-end data between endpoints without having to pass through each individual hop. PAM is the method that is currently used throughout the network for end-to-end signaling.

Figure 8-19. ISUP End-to-End Signaling

[View full size image]



ISDN Signaling Indicators in the IAM

The following set of fields in the IAM *FCI* comprises what is known as the *Protocol Control Indicator (PCI)*:

- End-to-end method indicator
- Interworking indicator
- IAM segmentation indicator
- ISDN User Part indicator

These fields provide information about the protocol communication across the ISUP connection. The Protocol Control Indicator fields are of particular importance to ISDN because they identify whether ISDN signaling can be exchanged across the network. If the Interworking Indicator is set to *interworking encountered*, it indicates that a non-SS7 connection (such as MF signaling) has been used in a circuit connection. It also indicates that SS7 signaling cannot be exchanged across this connection because it would prevent an ISDN terminal from being able to relay signaling across the network that depended on an SS7 connection all the way.

The ISDN User Part indicator field indicates whether ISUP has been used for every call leg up to the current exchange. If this field is set to *ISDN User Part not used all the way*, it might not be possible to pass ISDN information across the network.

The ISDN User Part preference indicator field indicates to the receiving node whether the call needs an outgoing ISUP connection.

The preference field might contain the following values:

- ISDN User Part preferred
- ISDN User Part required
- ISDN User Part not required

For calls originating from an ISDN set, the preference field is set to *ISDN User Part preferred* unless specified otherwise by different services. If it is available during outgoing trunk selection, an ISUP facility is chosen; an ISUP facility is "preferred," but not necessarily required. If an ISUP facility is not available, the call is still set up if a non-ISUP facility is available. If a call is being established that requires the ability to pass service information—such as end-to-end signaling—across the network, the preference field is set to *ISDN User Part required*. A call with a preference of "required" is not set up unless an ISUP facility is available. For example, setting up a multichannel ISDN video connection would not be possible without end-to-end ISUP signaling.

Although the PCI provides information about the connection across the network, it does not specify the actual protocol of the access signaling. The FCI includes the *ISDN access indicator* bit to indicate whether the originating terminal is an ISDN set.

Supplementary Services

Supplementary services are one of the ISUP advantages noted in this chapter's introduction. ISUP provides many messages and parameters that are explicitly created for the support of supplementary services across the network. The introduction of ISUP has helped to greatly standardize widely used services, allowing them to operate across networks and between vendors more easily. Service specifications still vary between different networks based on differences in locales and market needs. ISUP provides the flexibility to accommodate these differences using a rich message set and a large set of optional parameters.

The ITU-T defines a core set of widely used ISDN services in the Q.730–Q.739 series of specifications using ISUP network signaling. The actual specification of these services at the national level can vary. In addition, national networks and private networks offer many services outside of those that are specified by the ITU-T. In the United States, Telcordia has defined a large number of services in various Generic Requirements (GR) specifications for U.S. network operators.

The list of services implemented on modern telephony switches has grown quite long. However, the purpose of this section is not to explore the services themselves, but to provide examples of how ISUP is used to support them. Two examples of common services have been chosen to discuss how ISUP provides support for them: Calling Line Identification and Call Forwarding Unconditional.

Calling Line Identification (CLI) Example

ITU Q.731 specifies Calling Line Identification (CLI). Calling party information can be used at the terminating side of a call in many different ways. Following are a few examples:

- Calling Number Delivery (CND)
- Calling Name Delivery (CNAMD)
- Incoming Call Screening
- Customer Account Information Retrieval (Screen Pops)

Being able to identify the calling party allows the called party to make decisions before answering a call. For example, an end user can use call screening to allow them to choose which calls they wish to accept. A business might use the incoming number to speed the retrieval of customer account information to call centers. If the called party subscribes to Calling Name Delivery, the CgPN is used at the terminating exchange to retrieve the name associated with the number.

CLI is specifically defined by the ITU-T as:

- Calling Line Identification Presentation (CLIP)
- Calling Line Identification Restriction (CLIR)

The ISUP CdPN parameter contains an *Address Presentation Restricted indicator* that specifies whether the calling party identification can be presented to the called party. The Address Presentation Restricted indicator has the following possible values:

- Presentation allowed
- Presentation restricted
- Address not available
- Reserved for restriction by the network

If the terminating party subscribes to the CLI service, the terminating exchange uses this indicator's value to determine whether the number can be delivered. The number is delivered only if the value is set to Presentation allowed. If the connection encounters non-SS7 interworking, the address information might not be available for presentation. In addition, transit network operators might not transport the information in some cases, depending on regulatory policies. While the actual display to the end-user varies depending on location, it is quite common to see restricted addresses displayed as "private" and unavailable addresses displayed as "unknown" or "out of area."

In some networks, if the CLI is not present in the IAM, it might be requested from the calling party using an Information Request (INR) message. The originating exchange delivers the requested CLI using an Information (INF) message.

Call Forwarding Example

Call Forwarding is part of a larger suite of services known as Call Diversion services. There are many variations of Call Forwarding. The ITU-T in the Q.732 specification defines the standard set of Call Forwarding variations as follows:

- Call Forward Unconditional (CFU)
- Call Forward No Reply (CFNR)
- Call Forward Busy (CFB)

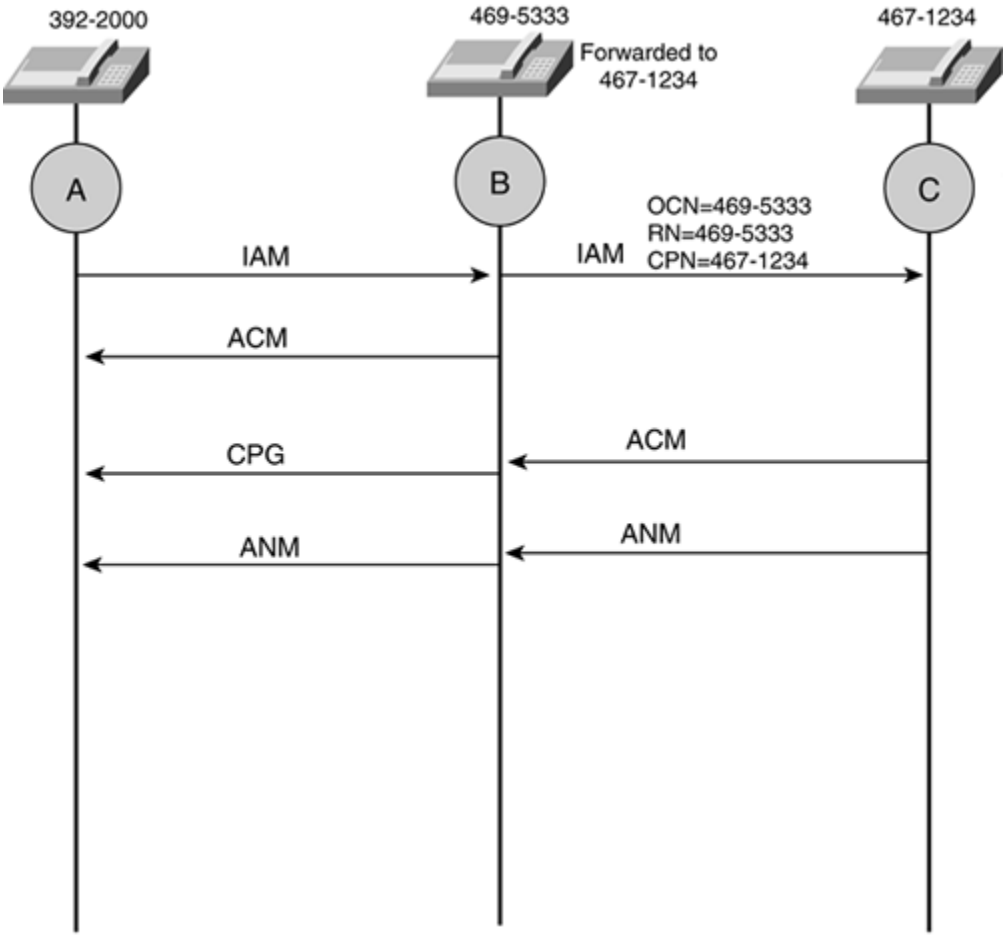
Other variations of Call Forwarding exist within localized markets. For example, Call Forwarding Selective is another variation that allows forwarding for calls that originate from

selective calling numbers. For this example, we have chosen Call Forward Unconditional to illustrate the use of ISUP signaling.

In Figure 8-20, the ITU-T message flow is shown for CFU at SSP B. The ANSI message flow differs slightly from that shown for ITU. A subscriber at SSP B has forwarded their calls to a number at SSP C. When SSP B attempts to terminate the call and encounters the Call Forward service, a new IAM is sent to SSP C. Keep in mind that a call might be forwarded multiple times before reaching its destination. The additional parameters included in the IAM for Call Forwarding convey information about the first and last instances of forwarding. In our example, the IAM to SSP C contains the following parameters, specific to the call redirection:

- Redirection Information (RI)
- Redirecting Number (RN)
- Original Called Number (OCN)

Figure 8-20. ISUP Call Forwarding Signaling



The inclusion of the RI parameter varies among different networks, so it might or might not be present. The RI parameter contains the following information fields:

- **Redirecting Indicator**— *Not specified for ANSI networks*. This field indicates how the call was forwarded and the presentation restriction indicators regarding the RI and RN.
- **Original Redirecting Reason**— Indicates why the first forwarding station forwarded the call (for example, *no reply* or *unconditional*). This field is set to *unconditional* in the example illustrated in Figure 8-20.
- **Redirection Counter**— Indicates the number of times a call has been forwarded. This counter is used to eliminate forwarding loops where a call ties up network resources because it is forwarded an excessive number of times. The ITU and ANSI standard for maximum redirections is five. In ANSI networks, the Hop Counter parameter provides this counter when RI is not included for forwarded calls. This field is set to 1 in the example illustrated in Figure 8-20.
- **Redirecting Reason**— Indicates the reason the call is being forwarded. In our example using CFU, the reason indicator is set to *unconditional*.

The OCN is the number dialed by the originator at A. The RN is the number of the station that forwarded the call. The RN is usually the same as the OCN, unless the call has been forwarded multiple times. If multiple forwarding have occurred, the RN is the number of the last station that forwarded the call. The CdPN will be set to the "forwarded to" number. Translation and routing using the new CdPN from the forwarding service at SSP B determine that the call should be directed to SSP C.

At SSP B, an ACM is returned to the originator and a new call is attempted to the forwarding destination. *Note that for ANSI networks, an ACM is not returned until the ACM is received from the new destination exchange, therefore, eliminating the CPG message.*

Additional Call Processing Messages

In addition to messages presented in the chapter, many other messages are used in various contexts for call processing. Some of the additional messages are used to support supplementary services, while others indicate specific network actions. Appendix B, "ISUP Messages (ANSI/UK/ETSI/ITU-T)," includes a complete list of all ISUP messages, their binary encoding, and a brief description.

Maintenance Messages and Procedures

ISUP provides an entire category of messages that are commonly categorized as "maintenance" messages. Until now, this chapter has focused on the call processing aspect of ISUP. This section discusses those messages that are used for diagnostics, maintenance, and the manipulation of ISUP facilities outside of the normal call processing realm.

The exchange can autonomously generate some maintenance messages, such as blocking (BLO) and Continuity Check Request (CCR), in response to an event or invoked manually by maintenance personnel. The collective set of messages described here helps to maintain trunk facilities and the integrity of user traffic. When necessary, trunks can be blocked from user traffic, tested, and reset to a state of sanity. The following sections illustrate how ISUP maintenance is used to accomplish these tasks:

- Circuit Ranges
- Circuit States

- Circuit Validation
- Continuity
- Blocking and Unblocking Circuits
- Circuit Reset

Circuit Ranges

ISUP maintenance messages apply to the CIC that is designated in the ISUP message. However, many messages can be applied to a range of CICs. These messages are referred to as "group" messages. Since ISUP trunk circuits are usually multiplexed together on digital spans, an action must often be applied to a larger group of circuits, such as the entire span. If a span is removed from service or brought into service, ISUP messages are sent to update the status of each of the span's circuits. If multiple spans are involved and individual messages were sent for each circuit, a flood of messages would occur over the SS7 network. Not only does this consume additional bandwidth on the SS7 links, but it also requires more processing by both the sending and receiving nodes. Using a single message with a CIC range eliminates the need to send a message for each CIC. Blocking messages, which we discuss later in this section, are a good example of where ranges are often used.

It is important to be aware that a message range can only be sent for contiguous CICs. If a span's CIC ranges were numbered using only even numbers such as 0, 2, 4,and 6, a message with a CIC range could not be used; individual messages would have to be sent for each CIC. It is good practice to number a span's CICs contiguously to maximize the efficiency of CIC ranges and effectively minimize message traffic.

Circuit States

An exchange maintains a circuit state for each bearer channel. Maintenance procedures and messages can affect that state. For example, maintenance messages can be sent to make circuits available for call processing, remove them from service, or reset them. A trunk circuit can have one of the following states:

- **Unequipped**— Circuit is not available for call processing.
- **Transient**— Circuit is waiting for an event to occur in order to complete a state transition. For example, an REL message has been sent, but an RLC has not been received.
- **Active**— Circuit is available for call processing. The circuit can have a substate of idle, incoming busy, or outgoing busy.
- **Locally blocked**— the local exchange has initiated the blocking of the circuit.
- **Remotely blocked**— the remote exchange has initiated the blocking of the circuit.
- **Locally and remotely blocked**— both the local and remote exchanges have initiated blocking.

The following messages are used for querying the state of a group of circuits. These messages are usually sent in response to maintenance commands entered at a maintenance interface, or by automated trunk diagnostics that are performed as part of routine trunk testing.

- **Circuit Query Message (CQM)**— Sent to the far end exchange to query the state of a group of circuits. This allows the states to be compared to ensure that the two nodes agree on the status of the facilities. It provides a safeguard against a state

mismatch in the event that a message indicating a change of state is sent, but not received.
- **Circuit Query Response Message (CQR)**— Sent in response to a CQM to report the state of the requested group of circuits.

Circuit Validation (ANSI Only)

Circuit validation determines whether translations data specific to the selection of an ISUP circuit has been set up correctly. The translations data at both ends of a circuit and between two exchanges is verified to ensure that the physical bearer channel can be derived. All switching systems require provisioning data to create the proper associations between trunkgroups, trunk members, CICs, and physical trunk circuits. Circuit Validation testing traverses these associations to ensure that they have been properly created. The Circuit Validation Test is particularly useful when turning up new trunk circuits because there is a greater potential for errors in newly provisioned facilities.

The Circuit Validation Test is typically invoked through a user interface at the switching system. Translations data at the local end is verified before sending a CVT message to the far end. The following messages are exchanged to perform the test:

- **Circuit Validation Test (CVT)**— Sent to the far end exchange to validate circuit-related translations data for an ISUP circuit. *This message is only used in ANSI networks*.
- **Circuit Validation Response (CVR)**— Sent in response to a CVT message to report the results of a Circuit Validation Test. The CVR message reports a success or failure for the Circuit Validation Test, along with characteristics of the circuit group being tested. For example, one reported characteristic is the method of glare handling being used for the circuit group. *This message is only used in ANSI networks*.

Continuity Testing

We have discussed continuity testing in the context of call processing where a circuit is tested before setting up a call. Continuity testing can also be performed manually by maintenance personnel, or by automated facilities testing.

The maintenance test procedure is slightly different than when it is performed as part of call processing. You will recall from the section on continuity testing that an indicator in the IAM is used for signifying that a test is required. When invoked as part of a maintenance procedure, the Continuity Check Request (CCR) message is used to indicate that a continuity test is required. The CCR is sent to the far end, and the continuity test proceeds as we discussed previously. The far end sends back a Loop Back Acknowledgement to acknowledge that a loop back or transceiver circuit has been connected for the test. The results are reported using a COT message by the node that originated the test. For additional information on continuity testing, refer to the "Continuity Test" section of this chapter. The following messages are used during the maintenance initiated continuity test:

- **Continuity Check Request (CCR)**— Sent to the far end to indicate that a continuity test is being performed. The far end connects a loopback or transceiver for the test.
- **Loop Around (LPA)**— Sent in response to a CCR to indicate that a loop back or transceiver has been connected to a circuit for continuity testing.

- **Continuity Test (COT)**— Sent to the far end to report the results of the continuity test. Indicates success if the received COT tones are within the specified guidelines of the country's standards. Otherwise, the message indicates a failure.

Blocking and Unblocking Circuits

ISUP provides blocking to prevent call traffic from being sent over a circuit. Maintenance messages can continue to be sent over the circuit. The two primary reasons for blocking are to remove a circuit from use when a problem has been encountered, or to allow for testing of the circuit. The local software blocks a trunk's local end. A blocking message notifies the trunk's far end about blocking. Unblocking is performed when circuits are ready to be returned to service for call traffic. The exchange unblocks locally and sends an unblocking message to the far end to provide notification of the state change. Both blocking and unblocking messages are acknowledged to ensure that both ends of the circuit remain in sync concerning the state of the trunk. The following messages are used in blocking and unblocking circuits:

- **Blocking (BLO)**— Sent to the far end to indicate the blocking of a circuit.
- **Blocking Acknowledgement (BLA)**— Sent as an acknowledgement in response to a BLO.
- **Circuit Group Blocking (CGB)**— Sent to the far end to indicate blocking for a range of circuits. The CICs must be contiguous for the group of circuits being blocked.
- **Circuit Group Blocking Acknowledgement (CGBA)**— Sent as an acknowledgement in response to a CGB.
- **Unblocking (UBL)**— Sent to the far end to indicate the unblocking of a blocked circuit.
- **Unblocking Acknowledgement (UBA)**— Sent as an acknowledgement in response to a UBL.
- **Circuit Group Unblocking (CGU)**— Sent to the far end to indicate unblocking for a range of blocked circuits.
- **Circuit Group Unblocking Acknowledgment (CGUA)**— Sent as an acknowledgement in response to a CGU.

Circuit Reset

A circuit is reset as an attempt to recover from an error condition or an unknown state. There are several reasons a circuit might need to be reset. Memory corruption or a mismatch of trunk states by the trunk's local and remote ends are examples of the need to reset a circuit. Calls are removed if they are active on the circuit that is being reset. A circuit reset reinitializes the local resources that are associated with the circuit and returns it to an idle state so it can be used again. Note that only group resets receive an acknowledgement from the far end; an individual reset does not. The following messages are associated with circuit resets:

- **Reset Circuit (RSC)**— Sent to the far end to indicate that the circuit is being reset to the idle state.
- **Group Reset Circuit (GRS)**— Sent to the far end to indicate that a contiguous group of CICs are being reset.
- **Group Reset Acknowledgement (GRA)**— Sent as an acknowledgement in response to GRS.

Summary

ISUP provides a rich network interface to call processing at an SSP. The increased bandwidth and protocol standardization allow a greater range of services that are able to interwork both within a network and across network boundaries. ISUP was designed to interface well with ISDN access signaling by providing event mapping and facilitating end-to-end user signaling. The protocol's use of optional message parameters achieves flexibility and extensibility.

ISUP uses a CIC identifier in each message to correlate the signaling with the correct circuit. The CIC is the key to associating signaling with bearer circuits.

ISUP also provides a set of maintenance messages for diagnostics and maintenance of ISUP facilities. These messages allow for blocking, testing, and resetting circuits and inquiring about circuit status.

Chapter 10. Transaction Capabilities Application Part (TCAP)

The *Transaction Capabilities Application Part (TCAP)* of the SS7 protocol allows services at network nodes to communicate with each other using an agreed-upon set of data elements. Prior to SS7, one of the problems with implementing switching services beyond the boundary of the local switch was the proprietary nature of the switches. The voice circuits also had very little bandwidth for signaling, so there was no room for transferring the necessary data associated with those services. Moving to a Common Channel Signaling (CCS) system with dedicated signaling bandwidth allows the transfer of a greater amount of service-related information. Coupling the standardization of data communication elements with the necessary bandwidth to transmit those elements creates the proper foundation for a rich service environment. To that end, TCAP provides a generic interface between services that is based on the concept of "components." Components comprise the instructions that service applications exchange at different nodes.

instructions

This chapter examines components and other details of the TCAP protocol, including the following:

- Overview of TCAP
- Message types
- Transactions
- Components
- Dialogue portion
- Message encoding
- Element structure
- Error handling
- ITU protocol message contents
- ANSI protocol message contents
- ANSI national operations

In trying to understand how TCAP works, the differences between ANSI TCAP (as presented in the ANSI T1.114) and ITU TCAP (as presented in the Q.700 series) are normalized as much as possible. While differences between the two certainly exist, a great deal of commonality also exists and often varies only in the naming of identifiers.

Overview

The following topics provide an overview of TCAP and how it is used to provide enhanced network services:

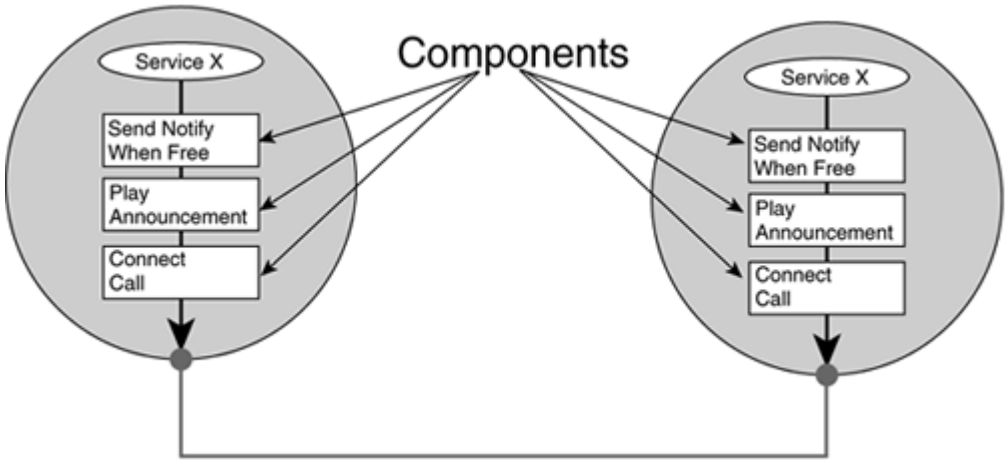
- Generic service interface
- Role of TCAP in call control
- TCAP within the SS7 protocol stack
- Transaction and component sublayers

Generic Service Interface

TCAP is designed to be generic to accommodate the needs of a wide variety of different services. This chapter focuses on understanding these generic mechanisms. Chapter 11, "Intelligent Networks (IN)," examines the prominent network services that use TCAP in an effort to understand how services use these generic mechanisms. Some common services that use TCAP include number translation services, such as Enhanced 800 Service (toll-free) and Local Number Portability (LNP). Other examples of TCAP users are Custom Local Area Signaling Services (CLASS), Mobile Wireless, and Advanced Intelligent Network (AIN) services. Figure 10-1 shows how TCAP uses standardized components as the basic building blocks for services across network nodes.

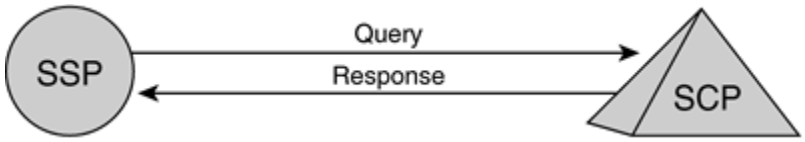
Figure 10-1. Standardized Components Used to Create a Generic Interface

[View full size image]



Most TCAP services can be viewed as a dialogue of questions and answers. A switch needs additional information that is associated with call processing, or with a particular service that causes it to send a TCAP query that requests the needed information. As shown in Figure 10-2, the answer returns in a TCAP response, which provides the necessary information, and normal call processing or feature processing can resume. The query for information can be sent to a Service Control Point (SCP) or to another SSP, depending on the type of service and the information required. The SCP is an SS7-capable database that provides a centralized point of information retrieval. It typically handles number translation services, such as toll-free and LPN; however, SCPs are also used for a number of additional IN/AIN services.

Figure 10-2. Simple Query and Response



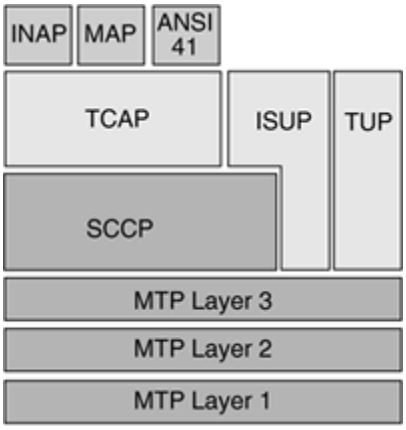
Role of TCAP in Call Control

TCAP is used to provide information to SSPs. This information is often used to enable successful call completion, but TCAP is not involved in the actual call-setup procedures. The protocol's circuit-related portion, such as ISUP and TUP, perform the call setup. This interaction between the service information provided by TCAP and the circuit-related protocol that performs the call setup occurs at the application level, not at the SS7 protocol layer. Within the SSP, the switching software that is responsible for call processing interacts with both the TCAP side of the SS7 stack and the call setup side of the stack (ISUP, TUP) to complete the call.

TCAP within the SS7 Protocol Stack

As shown in Figure 10-3, TCAP is at level 4 of the SS7 protocol stack. It depends upon the SCCP's transport services because TCAP itself does not contain any transport information. First, SCCP must establish communication between services before TCAP data can be delivered to the application layer. Refer to Chapter 9, "Signaling Connection Control Part (SCCP)," for more information on SCCP's transport services. TCAP interfaces to the application layer protocols above it, such as the ITU Intelligent Network Application Part (INAP), ANSI AIN, and ANSI-41 Mobile Switching to provide service-related information in a generic format. The application layer that passes information down to be encapsulated within TCAP is known as a Transaction Capability User (TC-User). The terms application, service, and TC-User are used interchangeably.

Figure 10-3. TCAP within the SS7 Stack



Transaction and Component Sublayers

The TCAP message is composed of two main sections: the transaction sublayer and the component sublayer. A transaction is a set of related TCAP messages that are exchanged between network nodes. The transaction portion identifies the messages that belong to the same transaction using a Transaction Identifier (TRID). The message's component portion contains the actual instructions, or "operations," that are being sent to the remote application. This chapter examines both areas in detail, along with the procedures surrounding their use.

Message Types

The TCAP *message type* (which is referred to as package type in ANSI) identifies the type of message being sent within the context of a transaction. Table 10-1 lists the seven package types for ANSI and Table 10-2 lists the five message types for ITU.

Table 10-1. Package Types for ANSI

ANSI Package Types	Hex Value	Description
Unidirectional	11100001	Sent in one direction and expects no reply.
Query with Permission	11100010	Initiates a transaction, giving the receiving node permission to end the transaction.
Query without Permission	11100011	Initiates a transaction but does not allow the receiving node to end the transaction
Response	11100100	Ends a transaction.
Conversation with Permission	11100101	Continues a transaction, giving the receiving node permission to end the transaction.
Conversation	11100110	Continues a transaction, but does not allow the receiving

Table 10-1. Package Types for ANSI

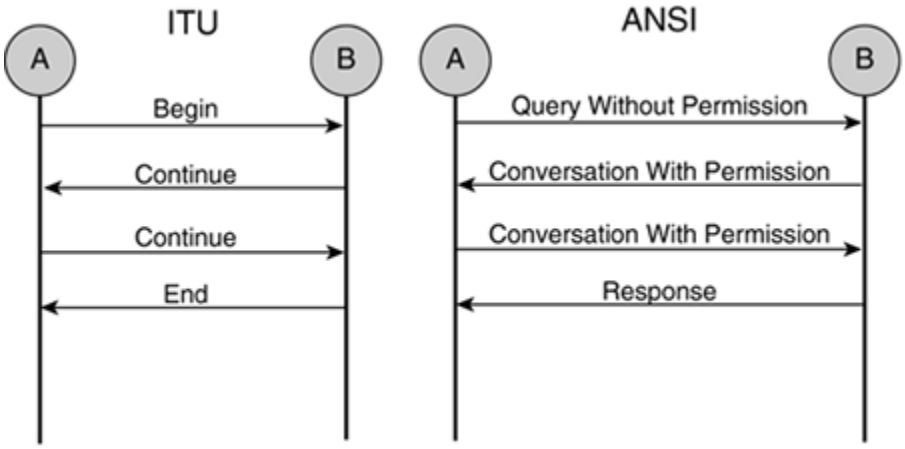
ANSI Package Types	Hex Value	Description
without Permission		node to end the transaction.
Abort	11110110	Sent to notify the destination node that an established transaction has been terminated without sending any further components that might be expected.

Table 10-2. Message Types for ITU

ITU Message Types	Hex Value	Description
Unidirectional	01100001	Sent in one direction and expects no reply.
Begin	01100010	Initiates a transaction.
(Reserved)	01100011	Not used.
End	01100100	Ends a transaction.
Continue	01100101	Continues an established transaction.
(Reserved)	01100110	Not used.
Abort	01100111	Sent to notify the destination node that an established transaction has been terminated without sending any further components that might be expected.

The message type also infers the stage of transaction processing. Figure 10-4 shows an example of an ITU conversation and an equivalent ANSI conversation. In ITU, a Begin message always starts a transaction, and an End message normally ends the transaction. (The "Transactions" section of this chapter discusses an exception to this rule.) The equivalent ANSI messages that begin and end transactions are Query (with or without permission) and Response, respectively. Conversation (ANSI) and Continue (ITU) messages indicate that further communication is required in an existing transaction.

Figure 10-4. Examples of ITU and ANSI Message Flow



Transactions

The services that use TCAP vary in complexity. Some require a node to translate and receive only a single message. For example, a basic toll-free call typically works in this manner. Other services, such as Call Completion to a Busy Subscriber (CCBS), can exchange a number of messages between nodes.

A *transaction* is a set of related messages that are exchanged between application processes at two different nodes. At any time, a node can have many simultaneous transactions in progress and send and receive multiple TCAP messages. For example, several subscribers might invoke a CCBS during the same period of time.

NOTE

CCBS is a subscriber feature used for completing calls to a busy subscriber by monitoring the called party's line and completing a call attempt when the called party is free. TCAP messages are exchanged between the telephony switches of the calling and called parties to monitor the busy line and provide notification when it is free. The service is also popularly known as *Automatic Callback*.

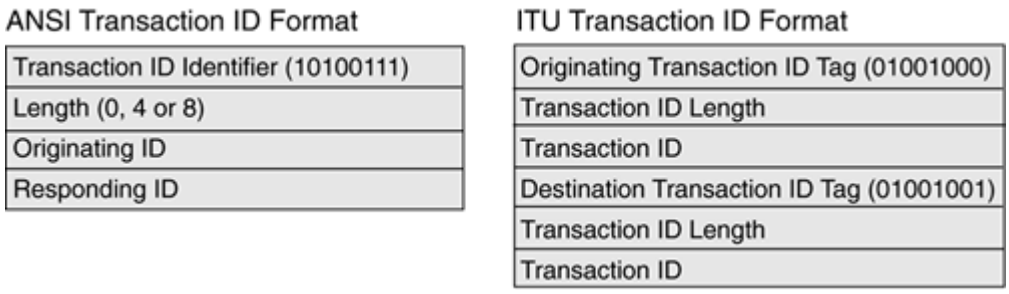
When a node sends a message and expects a reply, the sending node establishes and maintains a Transaction ID. This allows an incoming message to be properly associated with previously sent messages.

Transaction IDs

Transactions always begin with an initiating TCAP message that contains an *Originating Transaction ID*. When the service has completed, the Transaction ID becomes available for use again by the application. Each transaction must have a unique Transaction ID for all outstanding transactions. When an ID is in use, it cannot be used again until the current transaction releases it. If the same ID belonged to two transactions, the system that received a message would not know the transaction to which it belonged. The ANSI Transaction ID is 4 octets in length, thereby allowing a total number of 2^{32} concurrent transactions to exist at a given time. The ITU Transaction ID is variable from 1 to 4 octets. Up to two Transaction IDs can be included in a TCAP message, an Originating Transaction ID, and a Responding Transaction ID (called a Destination Transaction ID in ITU). ANSI

packages the Transaction IDs differently than ITU by nesting both IDs within a single Transaction ID Identifier. The following figure shows the Transaction ID section.

Figure 10-5. Transaction ID Format



Establishing Transaction IDs

The node that originates the transaction assigns an Originating Transaction ID, which the node sends to the destination in the first message, to establish the transaction. When the destination node receives a message, the application examines its contents and determines whether it should establish its own transaction.

When the destination node responds to the originating node, the message that is sent contains a Responding (or Destination) Transaction ID. The Responding Transaction ID is the same as the Originating Transaction ID that was received in the Begin/Query message. It can be thought of as a reflection of the Originating ID. The destination node examines the contents of the message to determine if it should establish a transaction with the originating node. If establishing a transaction is necessary, an Originating Transaction ID is assigned by the responding destination node and placed in an ANSI Conversation or ITU Continue message along with the Responding Transaction ID to be sent back to the transaction originator. In this situation, each node establishes a transaction from its own point of view. Depending on the message type, a TCAP message can contain zero, one, or two Transaction IDs. Tables 10-3 and 10-4 show the relationship between a message type and Transaction IDs for ITU and ANSI, respectively. For example, in Table 10-3, a Unidirectional message does not contain any Transaction IDs, while a Continue message contains two Transaction IDs.

Table 10-3. ITU Message Transaction IDs

ITU Message Type	Originating Transaction ID	Destination Transaction ID
Unidirectional	No	No
Begin	Yes	No
End	No	Yes
Continue	Yes	Yes
Abort	No	Yes

Table 10-4. ANSI Message Transaction IDs		
ANSI Package Type	**Originating Transaction ID**	**Responding Transaction ID**
Unidirectional	No	No
Query with Permission	Yes	No
Query without Permission	Yes	No
Response	No	Yes
Conversation with Permission	Yes	Yes
Conversation without Permission	Yes	Yes
Abort	No	Yes

Releasing Transaction IDs

The communicating applications can end transactions in one of two ways: either with a terminating message or a prearranged end. The most common method is a terminating message—a Response package in ANSI and an End message in ITU. The prearranged transaction end is simply an agreement at the application layer that a transaction ends at a given point. Releasing the Transaction ID returns it to the available pool of IDs so that another transaction can use it.

Transaction Message Sequence

Applications do not always establish a transaction during TCAP communications. The Unidirectional message is used to communicate when no reply is expected, therefore, requiring no Transaction ID. All other messages require a Transaction ID.

Figure 10-6 shows an example of a transaction occurring between two SS7 nodes. A conversation is established between Node A and Node B. As mentioned previously, a Query or Begin message always initiates a transaction. Node A establishes a transaction with a Transaction ID of 0. When the service logic at Node B processes the incoming message, it determines that it is necessary to establish a transaction from its own point of view. This is usually done to request additional information from the node that sent the message. In this example using the ANSI protocol, node B does not have a choice about engaging in a conversation because it has received a "Query without Permission" message. The message's "without Permission" designation is used to deny the receiving node the opportunity to end the transaction until it receives permission. In this example, Node B initiates a transaction with a Transaction ID of 1, thereby associating it with the received Transaction ID of 0.

Figure 10-6. Transaction Example Using ANSI Protocol

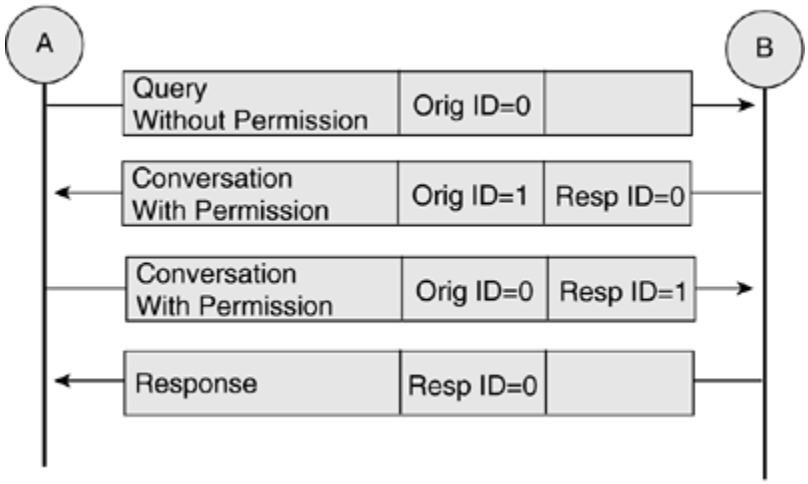
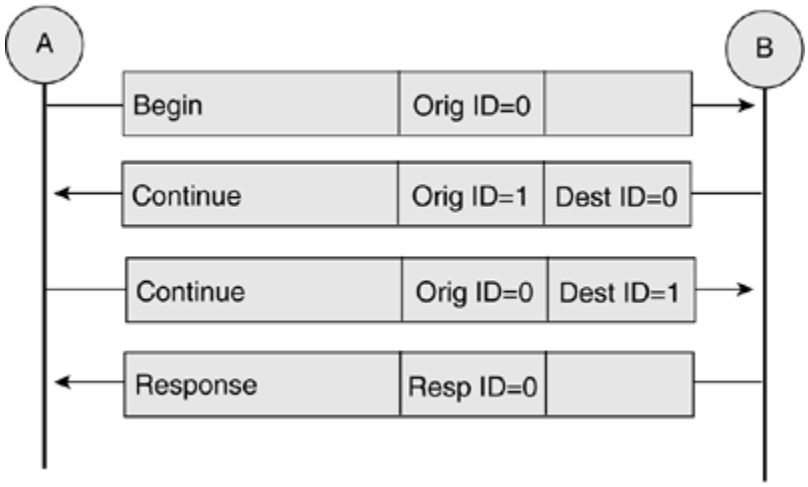


Figure 10-7 shows the same transaction using the ITU protocol. As shown by comparing the two examples, the two protocols are conceptually quite similar. Other than naming conventions and binary encoding, the primary difference is that the ITU message types do not explicitly state whether the receiving node must engage in a transaction from its perspective. This must be determined from the application logic.

Figure 10-7. Transaction Example Using ITU Protocol



Components

Components are a means of invoking an operation at a remote node. A TCAP message can contain several components, thereby invoking several operations simultaneously. The TCAP component is based on the ITU X.410 specification for Remote Operations in Message

Handling Systems. ITU X.229 has replaced this specification. The specification defines the following four Operational Protocol Data Units (OPDUs):

- **Invoke**— Requests an operation to be performed
- **Return Result**— Reports the successful completion of a requested operation
- **Return Error**— Reports the unsuccessful completion of a requested operation
- **Reject**— Reports a protocol violation, such as an incorrect or badly-formed OPDU

Each of the TCAP component types directly correlates to one of the previous OPDU types. The Invoke and Return Result component types are used for carrying out the normal operations between TCAP users. The Return Error and Reject component types are used for handling error conditions.

The contents of the Invoke and Return Result components include the following information:

- Component Type
- Component ID
- Operation Code (Invoke Component only)
- Parameters

The contents of the Return Error and Reject components are similar to the Invoke and Return Result components, except that the Operation Code used in an Invoke component is replaced by an Error/Problem code. The following sections discuss the contents of the components listed previously. The "Error Handling" section later in this chapter addresses the Return Error and Reject components.

Invoke and Return Result Components

Under normal circumstances, Invoke and Return Result Components are sent to carry out and verify operations between two communicating entities. For example, an SSP might "invoke" a number translation at an SCP, resulting in a new number being returned. A number of services, such as Toll-free, Premium Rate, and Local Number Portability, use TCAP to look up numbers in this manner. The application layer for these services and others use a standardized set of operations that is recognized by the network nodes involved in the communication. The information from the application layer is passed to the TCAP layer and encoded into components. Each Invoke Component is generally structured as an "instruction" and "data." The instructions are in the form of Operation Codes, which represent the operations that are being requested. The data is in the form of Parameters.

ITU Q.771 defines four classes of operations that determine how to handle Invoke replies. The TCAP message does not explicitly contain operation class information. Instead, it specifies the operation class using primitives between the application (TC-User) and the component sublayer.

NOTE

As used in this context, a primitive is a software indication that is used to pass information between software layers.

In other words, the indication of whether a reply is required and the tracking of whether that reply has been received are performed within the software. The main point is that operations can be handled differently, depending on the application logic. The four classes of operations are:

- **Class 1**— Success and failure are reported.
- **Class 2**— only failure is reported.
- **Class 3**— only success is reported.
- **Class 4**— neither success nor failure is reported.

The application logic is also responsible for determining whether an operation is a success or a failure. Based on the operation's results, a reply might be required. If a reply is required, one of the following components is sent:

- **Return Result**— Indicates a successfully invoked operation
- **Return Error**— Indicates a problem
- **Reject**— Indicates an inability to carry out an operation

Here we focus only on the Return Result component; the "Error Handling" section discusses the Return Error and Reject components. The following are the two types of Return Result components:

- Return Result Last
- Return Result Not Last

The Return Result Last indicates that an operation's final result has been returned. The Return Result Not Last indicates that further results will be returned. This allows the result to be segmented across multiple components.

ANSI TCAP also allows the use of an Invoke to acknowledge a previous Invoke. Because ANSI allows an Invoke to be used in response to another Invoke where a Return Result would otherwise be used, the Invoke also has two types: *Invoke Last* and *Invoke Not Last*. There is only a single Invoke type in ITU networks, and it is the equivalent of the ANSI Invoke Last component type.

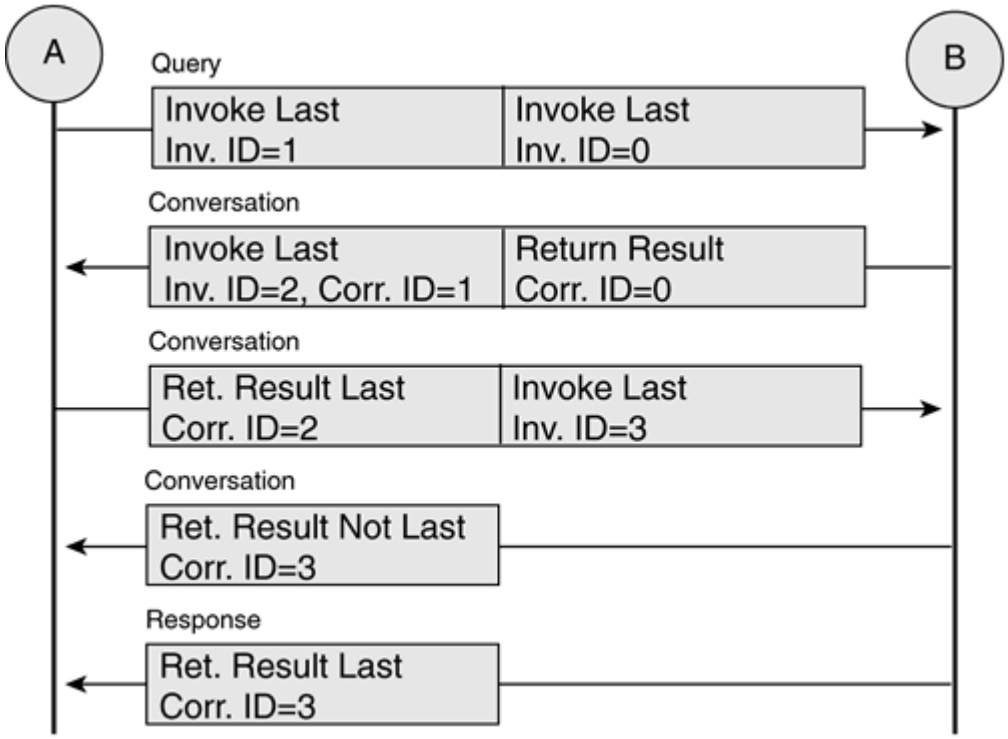
The details of segmenting results using the "Not Last" designation for both the Return Result and Invoke (ANSI Only) component types are more easily understood after a discussion of component IDs. We revisit this topic in a later section, after introducing correlation and linked-component IDs.

Component IDs

As mentioned previously, a message can contain several components. Each Invoke Component is coded with a numeric Invoke ID, which must be unique for each operation in progress because the ID is used to correlate the exchange of components for a particular operation. Just as a message can have several components, an operation can also have several parameters associated with it. Figure 10-8 shows an example of how Component

IDs are used in an ANSI network message exchange. Node A sends a message to Node B that contains two Invoke Components indicating that two remote operations are being requested. Node B processes the incoming components, carries out the requested operations, and sends an Invoke Component and a Return Result Component back to Node A. The Invoke component contains two IDs: an Invoke ID and a Correlation ID (linked ID in ITU-T networks). As shown in this example, an Invoke ID can be used to respond to another Invoke ID, rather than using a Return Result. Node B is requesting an operation from Node A using Invoke ID 2 in response to the previously received Invoke, reflecting ID 1 in the Correlation ID. The Return Result Component in the message contains a Correlation ID of 0 to reflect the previous Invoke with a Component ID of 0 from Node A. Node A then replies to the Invoke ID 2 with a Return Result and also invokes another operation using Invoke ID 3 in the same Conversation message. Finally, Node B answers with a Return Result Not Last Component for Invoke ID 3, followed by a Return Result Last for the same Component ID. This completes the component exchange between the communicating nodes. Notice that for each Invoke, a reply was received using either another Invoke with a "Reflecting" ID (the correlation or linked ID) or a Return Result (Last) Component. The Correlation ID shown in the figure is used as the "Reflecting" ID in ANSI networks; for ITU networks, the Linked ID serves as the "Reflecting" ID.

Figure 10-8. Component ID Association (ANSI Protocol)



Operation Codes

The *Operation Code* identifies the operation to be invoked at the node that receives the message. Operation Codes are context-dependent, meaning that they are understood only within the context of a particular service. For example, consider a caller who dials a toll-free number that is not supported in the region from which the caller is dialing. The SCP sends an Operation Code to the SSP for "Play an Announcement," instructing it to connect the subscriber to an announcement machine. The component that contains the "Play Announcement" Operation Code contains a parameter for identifying the proper announcement to be played. In this case, the caller hears an announcement that is similar to "The number you have dialed is not supported in your area."

ANSI defines a number of national Operation Codes in the ANSI TCAP specifications. In ITU networks, these definitions are typically relegated to layers above the TCAP protocol, such as INAP. Examples of these can be found in Chapter 11.

Parameters

Components can have *parameters* associated with them. The parameters are the data that is necessary to carry out the operation requested by the component Operation Code. For example, a component containing a "Play Announcement" Operation Code also contains an announcement parameter. The announcement parameter typically provides the announcement ID so the correct recording is played to the listener. Just as a TCAP message can contain multiple components, a component can contain multiple parameters.

The ITU-T does not define any specific parameters. This responsibility is relegated to national or regional standards bodies, such as ANSI and ETSI. Parameters can be defined as part of the TCAP standards (for example, ANSI standards) or relegated to the definition of the protocol layers above TCAP, such as INAP (for example, ETSI standards). ANSI defines a number of national parameters in the ANSI T1.114 specification. Application processes can use these parameters directly.

Chapter 11, "Intelligent Networks" provides examples of TCAP parameters that are defined by protocols above the TCAP layer. The AIN and INAP parameters described here are used in TCAP messages for ANSI and ITU-T networks, respectively.

ANSI parameters are specified either as part of a set or a sequence. A parameter set is used when parameters are delivered with no particular order and can be processed in any order.

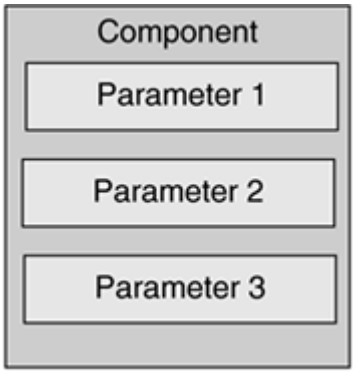
A parameter sequence specifies that the parameters should be processed in the order in which they are received.

ITU-T does not use parameter sequencing, so there is no designation of set or sequence. Parameters are handled in the same manner as an ANSI parameter set, with delivery occurring in no particular order.

ITU Parameters

Figure 10-9 shows a component with multiple ITU parameters.

Figure 10-9. Component with Multiple ITU Parameters



Dialogue Portion

The dialogue portion of the message is optional and is used to convey information about a dialogue between nodes at the component sublayer. It establishes a flow of information within a particular context for a transaction. Information, such as the protocol version and application context, is used to ensure that two nodes interpret the component sublayer's contents in the same manner using an agreed upon set of element definitions.

ITU Dialogue

There are two categories of dialogues: structured and unstructured. An unstructured dialogue is one in which no reply is expected. This type of dialogue uses a Unidirectional message type at the transaction layer. A structured dialogue requires a reply.

Within these two general categories of dialogues, dialogue-control Application Protocol Data Units (APDU) are used to convey dialogue information between TC-Users. The following are four types of APDU:

- Dialogue Request
- Dialogue Response
- Dialogue Abort
- Dialogue Unidirectional

Following is a description of each of these APDU and the information elements contained therein. The ITU unstructured dialogue uses the following dialogue-control APDU:

- **Unidirectional Dialogue**— The Unidirectional Dialogue consists of an Application Context Name and optional Protocol Version and User Information. It is used to convey dialogue information in one direction, for which no reply is expected.

The structured dialogue uses the following dialogue-control APDUs:

- **Dialogue Request**— The Dialogue Request consists of an Application Context Name and, optionally, Protocol Version and User Information. It is used to request dialogue information from another node, such as the context between the nodes (what set of

operations will be included) and to distinguish that the correct protocol version is being used to interpret the information that is being sent.
- **Dialogue Response**— The Dialogue Response is sent as a reply to a Dialogue Request. In addition to the information elements of the Dialogue Request, it includes a Result field and a Result Source Diagnostic element. The result indicates whether the dialogue has been accepted. If a Rejection indication is returned, the dialogue does not continue. In cases where rejection occurs, the Result Diagnostic indicates why a dialogue is rejected.

As you can see from the descriptions, a number of the dialogue information elements are common across the dialogue APDU types. Following is a brief description of the dialogue information elements:

- **Application Context Name**— identifies the application to which the dialogue components apply.
- **Protocol Version**— Indicates the version of the dialogue portion that can be supported. This helps ensure proper interpretation of the dialogue information between TC-Users when new versions of the dialogue portion are created.
- **User Information**— Information exchanged between TC-Users that is defined by and relevant only to the application. The contents of the user information element are not standardized.
- **Result**— Provides the initiating TC-User with the result of the request to establish a dialogue.
- **Result Source Diagnostic**— Identifies the source of the *Result* element and provides additional diagnostic information.
- **Abort Source**— Identifies the source of an abnormal dialogue release. The source might be the TC-User or the dialogue portion of the message.
- **Dialogue Abort**— The Dialogue Abort is used to terminate a dialogue before it would normally be terminated. The Dialogue Abort contains only an *Abort Source* and, optionally, *User Information*. The Abort Source is used to indicate where the Abort was initiated—from the user or the service provider.

ANSI Dialogue

The *ANSI Dialogue* can contain any of the following optional Dialogue elements. Note that the Application Context and Security can use either an integer for identification or an OID (Object Identifier). The OID is a common structure used for identifying objects in communications protocols by using a hierarchical tree notation such as "3.2.4."

- **Dialogue Portion Identifier**— This identifier indicates the beginning of the dialogue portion of the message. The following elements are included within this dialogue section.
- **Protocol Version**— Identifies the version of TCAP to be used in interpreting the message; for example, T1.114 version 1992 versus TCAP T1.114 version 1996.
- **Application Context Integer/Application Context OID**— Identifies the context in which to interpret the message. Since TCAP is generic and the operations must always be interpreted in the context of a particular service or set of services that use unique identifiers for each operation, this can be used to specify the context.
- **User Information**– —Provides additional information that is only relevant to the application, to assist the receiving TC-User (such as an application) in interpreting the received TCAP data. An example is including a version number for the application that uses the encapsulated TCAP components.

- **Security Context Integer/Security Context OID**— Used for establishing a secure dialog. The Security Context is used to determine how other security information, such as Confidentiality, should be interpreted.
- **Confidentiality Integer**— Used to specify how confidentiality is accomplished by providing encryption/decryption procedures. It contains the following optional fields. If neither of these optional fields is included, the confidentiality information is not used because no specification exists regarding how information should be protected or interpreted.

  - **Confidentiality Algorithm ID**— An integer or OID that identifies the algorithm to use for decoding encrypted data.

  - **Confidentiality Value**— Any information that can be encoded using Basic Encoding Rules (BER). The BER are the ITU X.690 ASN.1 (Abstract Syntax Notation) rules for encoding information into binary format for transmission.

Message Encoding

The TCAP data element encoding is based on the ITU X.680 and X.690 ASN.1 standards. Many of the SS7 standards reference the older versions of these documents (X.208 and X.209). The ASN.1 provides a means of describing complex data structures in a logical, readable text form and specifying encoding procedures for transmission in binary form.

The following example shows the ASN.1 definition for the ANSI TCAP package type and is taken directly from the ANSI T1.114 specification.

Example 10-1. The ANSI Definition for the ANSI TCAP Package Type

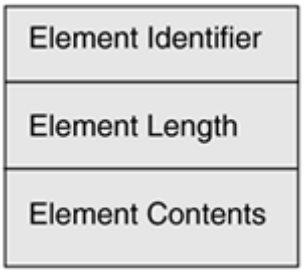
```
PackageType  ::= CHOICE { unidirectional  [PRIVATE 1]  IMPLICIT
UniTransactionPDU

                         QueryWithPerm     [PRIVATE 2]  IMPLICIT Transaction
PDU

                         queryWithoutPerm  [PRIVATE 3]  IMPLICIT Transaction
PDU

                         response          [PRIVATE 4]  IMPLICIT Transaction
PDU

                     conversationWithPerm [PRIVATE 5]  IMPLICIT Transaction
PDU

                 conversationWithoutPerm   [PRIVATE 6]  IMPLICIT Transaction
PDU

                         abort             [PRIVATE 22] IMPLICIT Abort }
```

The data is described in a precise way using textual description. In this example, the package type is a choice of one of the designated types—unidirectional, queryWithPerm, and so forth. Each is coded as a "Private" Class (which we discuss shortly) and has a defined numeric identifier. Also, the choice of the package type implies whether it is a UniTransactionPDU, a Transaction PDU, or an Abort. While this is a simple example, ASN.1 is used to describe very complex nested structures. You can find complete TCAP definitions in ASN.1 format in both the ANSI T1.114 and the ITU Q.773 specifications.

Element Structure

From a structural point of view, a TCAP message is a collection of data elements. Each element takes the form of Identifier, Length, and Contents. The TCAP element is the basic building block for constructing a message.

Figure 10-10. TCAP Element



The TCAP element is constructed with a commonly used data encoding scheme, which is often referred to as TLV: Tag, Length, Value format. The identifier specifies the type of element so that the receiving node can interpret its contents correctly. The length is the number of bytes in the element contents, beginning with the first byte past the element length. The contents are the actual data payload being transmitted.

Element Identifier

The *Element Identifier* is one or more octets comprised of bit fields that creates the class, form, and tag. Tables 10-5 and 10-6 list the values for the class and form. Bit H is the most significant bit.

Table 10-5. Class Values

Class	Bit Value Bits (HG)	Definition
Universal	00	Universal
Application-wide	01	International TCAP
Context-specific	10	Context Specific

Table 10-5. Class Values		
Class	**Bit Value Bits (HG)**	**Definition**
Private Use	11	National TCAP/Private TCAP

Table 10-6. Form Bit	
Form	**Bit F**
Primitive	0
Constructor	1

The class defines the identifier's scope or context. The universal class is used for identifiers that are defined in X.690 and do not depend on the application. Application-wide is used for international standardized TCAP. Context-specific identifiers are defined within the application for a limited context, such as a particular ASN.1 sequence. Private Use identifiers can be defined within a private network. These identifiers vary in scope and can represent elements within a national network, such as ANSI, or can be defined within a smaller private network. The tag bits (bits A-E) help to further determine whether the element is national or private. For more information, see "Identifier Tag" in this section.

Constructors and Primitives

The Form bit (Bit F) indicates whether the value is a primitive or constructor, as listed in Table 10-6. A primitive is simply an atomic value.

NOTE

An *atomic value* is one that cannot be broken down into further parts. Be careful not to confuse the term primitive, used here, with software primitives, used earlier in the chapter.

A constructor can contain one or more elements, thereby creating a nested structure. For example, a Component Tag is a constructor because a component is made up of several elements, such as the Invoke ID and Operation Code.

Identifier Tag

Bits A-E of the element identifier uniquely identify the element within a given class. If all bits are set to 1, this is a special indicator, which specifies that the identifier is octet-extended. In this case, one or more octets follow with additional identifier bits. This format

allows the protocol to scale in order to handle a potentially large number of identifiers. If Bit H in the extension octet is set to 1, the identifier is octet-extended to another octet. If it is set to 0, it indicates the identifier's last octet. In the following table, the identifier is extended to three octets using the extension mechanism. As previously noted, the identifier is further discriminated based on the tag bits. When coded as class Private Use, bits A-E are used for national TCAP. If bits A-E are all coded to 1, the G bit in the first extension octet (X13 in the example below) indicates whether it is private or national. The G bit is set to 0 for national or to 1 for private.

Table 10-7. Class Encoding Mechanism

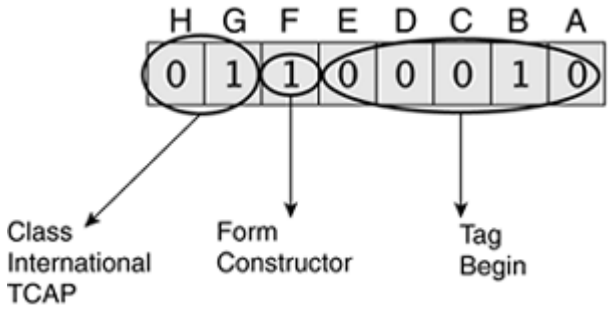
H	G	F	E	D	C	B	A	
CLASS		0	1	1	1	1	1	First Octet
1	X13	X12	X11	X10	X9	X8	X7	Second Octet
0	X6	X5	X4	X3	X2	X1	X0	Third Octet

An example illustrates how class, form, and tag are used to create a TCAP element. Figure 10-11 shows an ITU Begin message type in its binary form as it is transmitted on the signaling link. Bit A represents the least significant bit. The ITU Q.773 specification defines the ASN.1 description in the following manner:

Example 10-2. ASN.1 Definition for ITU Begin Message

```
MessageType ::= Choice {unidirectional [APPLICATION 1] IMPLICIT
Unidirectional,
                        Begin          [APPLICATION 2] IMPLICIT Begin,
```

Figure 10-11. ITU Begin Message Type Encoding



The message type is defined with a class of Application-wide and a tag of 2. It is a constructor because the message is comprised of multiple elements.

Length Identifier

The length field is also coded using an extension mechanism. If the length is 127 octets or less, Bit H is set to 0 and bits A-G contain the length. If the length is 128 or greater, Bit H is set to 1 and A-G contains the number of octets used to encode the Length field. The additional octets contain the actual length of the element contents. Table 10-8 shows an example using the extension mechanism to represent a length of 131 octets. The H bit is set in the first octet, and the value represented by bits A-G is 1; this means that one additional byte is used to represent the length. The second octet indicates that the element is 131 octets in length using standard binary representation.

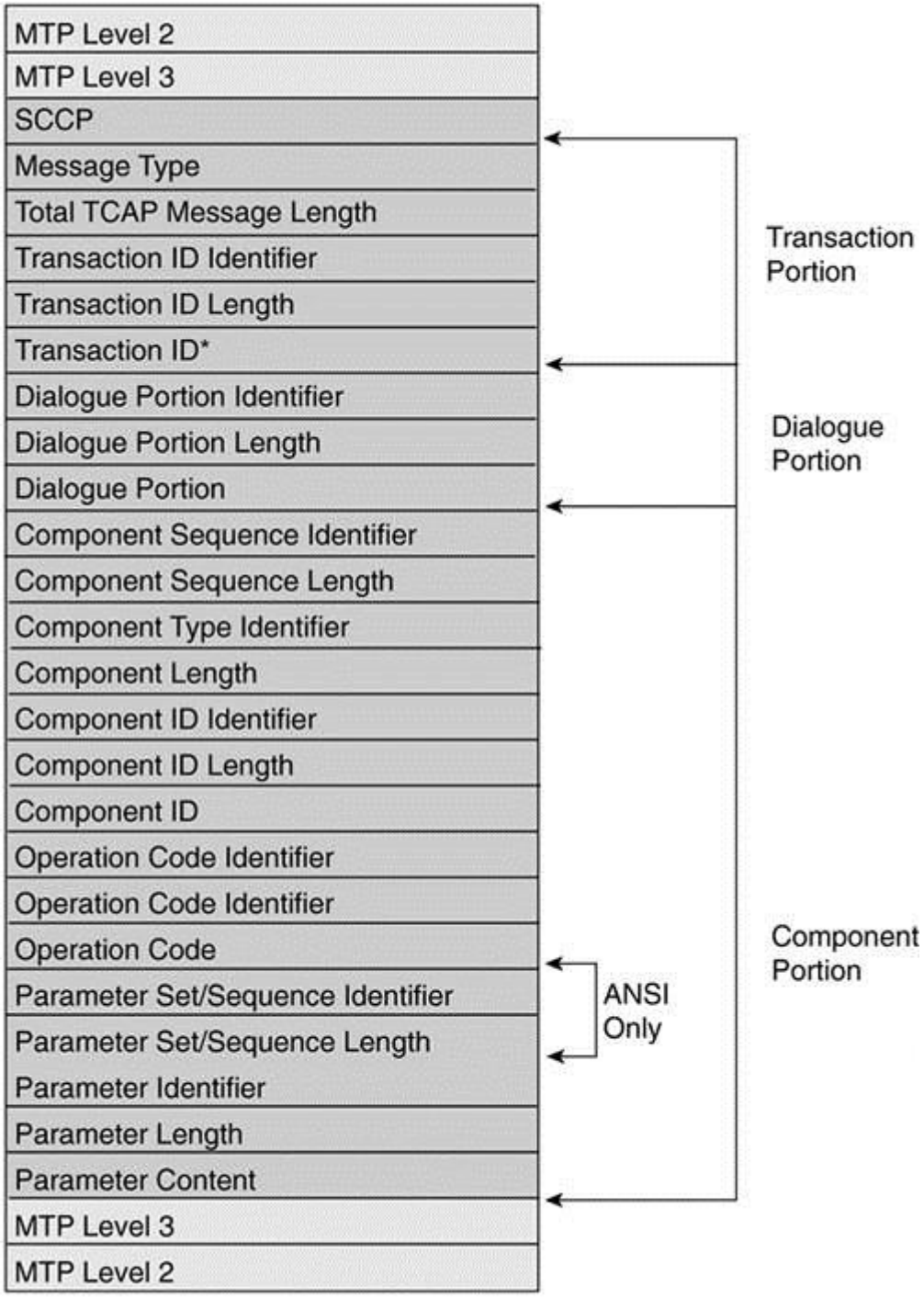
Table 10-8. Length Identifier Bits

Length Identifier Bits								
H	**G**	**F**	**E**	**D**	**C**	**B**	**A**	
1	0	0	0	0	0	0	1	First Octet
1	0	0	0	0	0	1	1	Second Octet

Message Layout

Now that we have examined in detail how each of the TCAP data elements are constructed, let's take a look at how they are assimilated into a message. There are three distinct sections into which a message is divided: the transaction, dialogue, and component portions. The dialogue portion of the message is optional. Figure 10-12 shows the complete structure of a TCAP message within the context of its supporting SS7 levels.

Figure 10-12. TCAP Message Structure



From the message structure, you can see the TLV format that is repeated throughout, in the form of Identifier, Length, and Content. A single component is shown with a single parameter in its parameter set; however, multiple parameters could exist within the component. If multiple parameters are included, another parameter identifier would immediately follow the Parameter Content field of the previous parameter. The message could also contain multiple components, in which case the next component would follow the

last parameter of the previous component. Only the maximum MTP message length limits the TCAP message length.

Error Handling

As with any other protocol, errors can occur during TCAP communications. TCAP errors fall into three general categories:

- Protocol Errors
- Application Errors
- End-user Errors

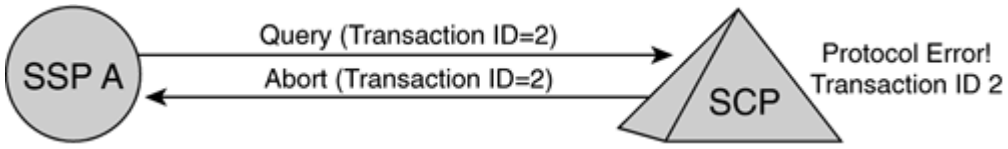
Protocol Errors

Protocol Errors are the result of TCAP messages being incorrectly formed, containing illegal values, or being received when not expected. For example, receiving an unrecognized message type or component type would constitute a protocol error. Another example of an error would be receiving a responding Transaction ID for a nonexistent transaction. While the actual value of the ID might be within the acceptable range of values, the lack of a transaction with which to associate the response causes a protocol error.

Errors at the Transaction Layer

Protocol Errors that occur at the transaction sublayer are reported to the remote node by sending an Abort message type with a P-Abort cause—in other words, a Protocol Abort. The Abort message is sent only when a transaction must be closed and a Transaction ID can be derived from the message in which the error occurred. Figure 10-13 shows an Abort message being sent for an open transaction in which a protocol error is detected.

Figure 10-13. Protocol Error Causes an Abort



Because no Transaction ID is associated with a Unidirectional message, no Abort message would be sent if the message was received with an error. If a Query or Begin message is received and the Originating Transaction ID cannot be determined because of the message error, the message is simply discarded and an Abort message is not returned to the sender.

If the Transaction ID can be determined, the Abort message is sent to report the error. Without the Transaction ID, there is no way for the sending node to handle the error because it cannot make an association with the appropriate transaction.

Errors at the Component Layer

Protocol errors at the component sublayer are reported using a Reject Component. The errored component's Component ID is reflected in the Reject Component. A number of different errors can be detected and reported. For example, a duplicate Invoke ID error indicates that an Invoke ID has been received for an operation that is already in progress. This results in an ambiguous reference because both operations have the same ID. Another type of error is a component that is simply coded with an incorrect value, such as an unrecognized component type. Refer to the TCAP specifications for a complete list of errors that can be detected and reported.

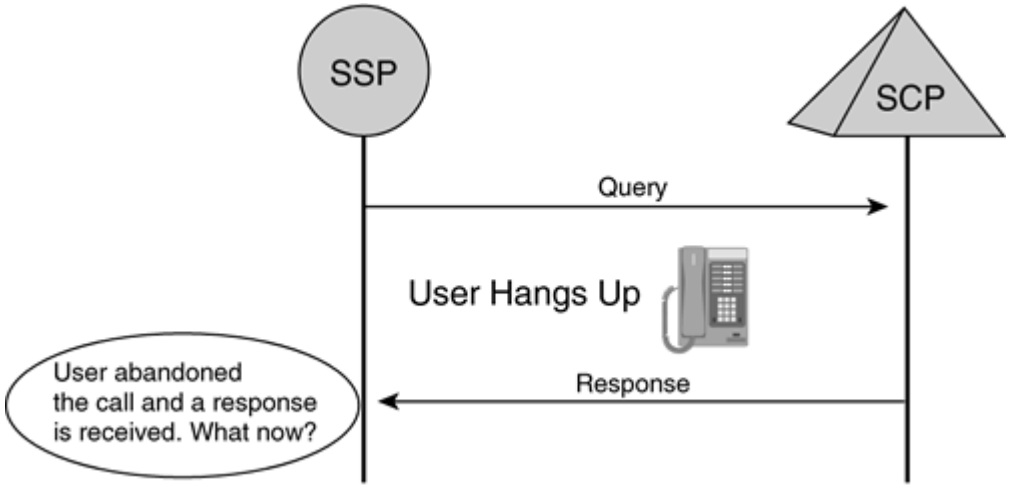
Application Errors

Application Errors are anomalies within the application procedure. An example is an unexpected component sequence, in which the received components do not match what the application procedures expect. Another example is a missing customer record error, which is an error that is used to indicate that a database lookup failed to find the requested information. The application is responsible for determining what actions to take when errors of this type are encountered.

End-User Errors

The *End-User Error* is similar to the Application Error in that it is an anomaly of the application procedure. However, as indicated by the name, the anomaly is the result of some variance from the normal actions by the user. The user might take an action, such as abandoning the call prematurely, as shown in Figure 10-14; or the user might enter an unexpected response when connected to a digit collection unit and prompted for input, thereby causing the error.

Figure 10-14. Error Caused by User Action



Handling Application and End-User Errors

Both the Application Error and the End-User Error are reported using the *Return Error* component for component-related errors. Because the errors in these two categories are actually variations in the application's script or procedure flow, the application determines how they are handled. These errors also imply that no error exists at the actual TCAP layer because a protocol error would trigger prior to an error at the application level. The application can also send an Abort message type (U-Abort) to the other node to indicate that a *User Abort* has occurred for the transaction and that it should be closed.

ITU Protocol Message Contents

The definition of each message type indicates a set of fields that comprise the message. While some fields are mandatory, others are optional. As specified by Q.773, the standard set of ITU messages includes:

- Unidirectional
- Begin
- End
- Continue
- Abort

The following sections describe these messages, the fields that are included in each one, and indicate which fields are mandatory or optional.

Unidirectional Message

The *Unidirectional Message* is sent when no reply is expected. Table 10-9 lists the message contents.

Table 10-9. Unidirectional Message Fields

Unidirectional Message Fields	Mandatory/Optional
Message Type Total Message Length	Mandatory
Dialogue Portion	Optional
Component Portion Tag Component Portion Length	Mandatory
One or More Componnts	Mandatory

Begin Message

The *Begin Message* is sent to initiate a transaction. Table 10-10 lists the message contents.

Table 10-10. Begin Message Fields

Begin Message Fields	Mandatory/Optional
Message Type Total Message Length	Mandatory
Originating Transaction ID Tag Transaction ID Length Transaction ID	Mandatory
Dialogue Portion	Optional
Component Portion Tag Component Portion Length	Optional[*]
One or More Components	Optional

[*] The component Portion Tag is present only if the message contains components.

EndMessage

The *End Message* is sent to end a transaction. Table 10-11 lists the message contents.

Table 10-11. End Message Fields

End Message Fields	Mandatory/Optional
Message Type Total Message Length	Mandatory
Destination Transaction ID Tag Transaction ID Length Transaction ID	Mandator
Dialogue Portion	Optional
Component Portion Tag	Optional[*]

Table 10-11. End Message Fields	
End Message Fields	**Mandatory/Optional**
Component Portion Length	
One or More Components	Optional

[*] The component Portion Tag is present only the message contains components.

Continue Message

The *Continue Message* is sent when a transaction has previously been established and additional information needs to be sent without ending the transaction. Table 10-12 lists the message contents.

Table 10-12. Continue Message Fields	
Continue Message Fields	**Mandatory/Optional**
Message Type Total Message Length	Mandatory
Originating Transaction ID Tag Transaction ID Length Transaction ID	Mandatory
Destination Transaction ID Tag Transaction ID Length Transaction ID	Mandatory
Dialogue Portion	Optional
Component Portion Tag Component Portion Length	Optional[*]
One or More Components	Optional

[*] The component Portion Tag is present only if the message contains components.

Abort Message

The *Abort Message* is sent to terminate a previously established transaction. Table 10-13 lists the message contents.

Table 10-13. Abort Message Fields

Abort Message Fields	Mandatory/Optional
Message Type Total Message Length	Mandatory
Destination Transaction ID Tag Transaction ID Length Transaction ID	Mandatory
P-Abort Cause Tag P-Abort Cause Length P-Abort Cause	Optional[*]
Dialogue Portion	Optional

[*] P-Abort is present when the TC-User generates the Abort message.

Summary

TCAP provides a standard mechanism for telephony services to exchange information across the network. It is designed to be generic so it can interface with a variety of services.

TCAP resides at Level 4 of the SS7 protocol and depends on SCCP's transport services. It is comprised of a transaction sublayer and a component sublayer. The transaction sublayer correlates the exchange of associated messages, while the component sublayer handles the remote operation requests.

All information elements in the TCAP message are defined and encoded using the syntax and BER of ASN.1. The ITU Q.771—Q.775 series of specifications defines the TCAP protocol. Specifications such as the ETSI.300.374 INAP series build on the ITU Q Series Recommendations to provide additional information needed for implementing network services. The ANSI T1.114 defines the TCAP specifications for ANSI networks. ANSI defines a number of national operations and parameters on which basic services can be built. Similar to ITU, many specifications build upon the basic TCAP provisions as defined in T1.114. For example, the Telcordia GR-1298 and GR-1299 AIN specifications provide the North American equivalent of the ETSI INAP service framework for IN services.

TCAP traffic on telephony signaling networks has increased in recent years because of an increase in services such as LNP, Calling Name Delivery, and Short Messaging Service (SMS), which rely on TCAP communication. This upward trend is likely to continue as IN services are more widely deployed, thereby making TCAP an increasingly important component in the role of network services.

Part III: Service-oriented Protocols

Chapter 11 Intelligent Networks (IN)

Chapter 12 Cellular Networks

Chapter 13 GSM and ANSI-41 Mobile Application Part (MAP)

Chapter 12. Cellular Networks

This chapter introduces Global System for Mobile communications (GSM), which is the most popular digital cellular network standard in terms of architecture, discusses interfaces and protocols, and concludes by presenting examples of mobility management and call processing in the network. The protocols that are found in GSM to perform these functions—namely, Base Station Subsystem Application Part (BSSAP) and Mobile Application Part (MAP)—are applications (subsystems) that utilize the underlying functionality of the SS7 protocols and network. This chapter aims to provide enough background on GSM cellular networks for you to understand the MAP that is used for mobility management and call processing within the GSM network, which is discussed in Chapter 13, "GSM and ANSI-41 Mobile Application Part (MAP)."

The European Telecommunication Standard Institute (ETSI) formulated GSM. Phase one of the GSM specifications was published in 1990, and the commercial operation using the 900 Mhz range began in 1991. The same year, a derivative of GSM, known as Digital Cellular System 1800 (DCS 1800), which translated GSM to the 1800 Mhz range, appeared. The United States adapted DCS 1800 into the 1900 Mhz range and called it Personal Communication System 1900 (PCS 1900). By 1993, 36 GSM networks existed in 22 countries [119].

Pre-GSM cellular networks are analog and vary from country to country—for example, the United States still uses Advanced/American Mobile Phone Service (AMPS), and the UK used Total Access Communication System (TACS). With these older analog standards, it was impossible to have one phone work in more than one country. In addition, because of the analog nature of the speech, quality could be relatively poor, and there were no provisions for supplementary services (such as call waiting). Although it is standardized in Europe, GSM is not just a European standard. At the time of this writing, there are more than 509 GSM networks (including DCS 1800 and PCS 1900) operating in 182 countries around the world, with 684.2 million subscribers [Source: GSM Association]. See Appendix I for a list of mobile networks by country.

GSM has been released in phases. The following are the features of these phases:

GSM Phase 1 (1992) Features

- Call Forwarding
- All Calls
- No Answer
- Engaged
- Unreachable
- Call Barring
- Outgoing—Bar certain outgoing calls
- Incoming—Bar certain incoming calls
- Global roaming—–If you visit any other country or parts in an existing country with GSM, your cellular phone remains connected without having to change your number or perform any action.

GSM Phase 2 (1995) Features

- **Short Message Service (SMS)**— Allows you to send and receive text messages.
- **Multiparty Calling**— Talk to five other parties and yourself at the same time.
- **Call Holding**— Place a call on hold.
- **Calling Line Identity Service**— This facility allows you to see the incoming caller's telephone number on your handset before answering.
- **Advice of Charge**— Allows you to keep track of call costs.
- **Cell Broadcast**— Allows you to subscribe to local news channels.
- **Mobile Terminating Fax**— Another number you are issued that can receive faxes.
- **Call Waiting**— Notifies you of another call while you are on a call.
- **Mobile Data Services**— Allows handsets to communicate with computers.
- **Mobile Fax Service**— Allows handsets to send, retrieve, and receive faxes.

GSM Phase 2 + (1996) Features

- Upgrades and improvements to existing services; the majority of the upgrade concerns data transmission, including bearer services and packet switched data at 64 kbps and above
- DECT access to GSM
- PMR/Public Access Mobile Radio (PAMR)-like capabilities to GSM in the local loop
- SIM enhancements
- Premium rate services
- Virtual Private Networks Packet Radio

Unlike Europe (and most of the world), which only pursued GSM for digital cellular networks, North America has pursued a mix of TDMA (IS-54, IS-136), CDMA, and GSM. At the time of this writing, TDMA and CDMA have been more widely deployed in North America than GSM. However, this situation is rapidly beginning to reverse with GSM continually gaining ground.

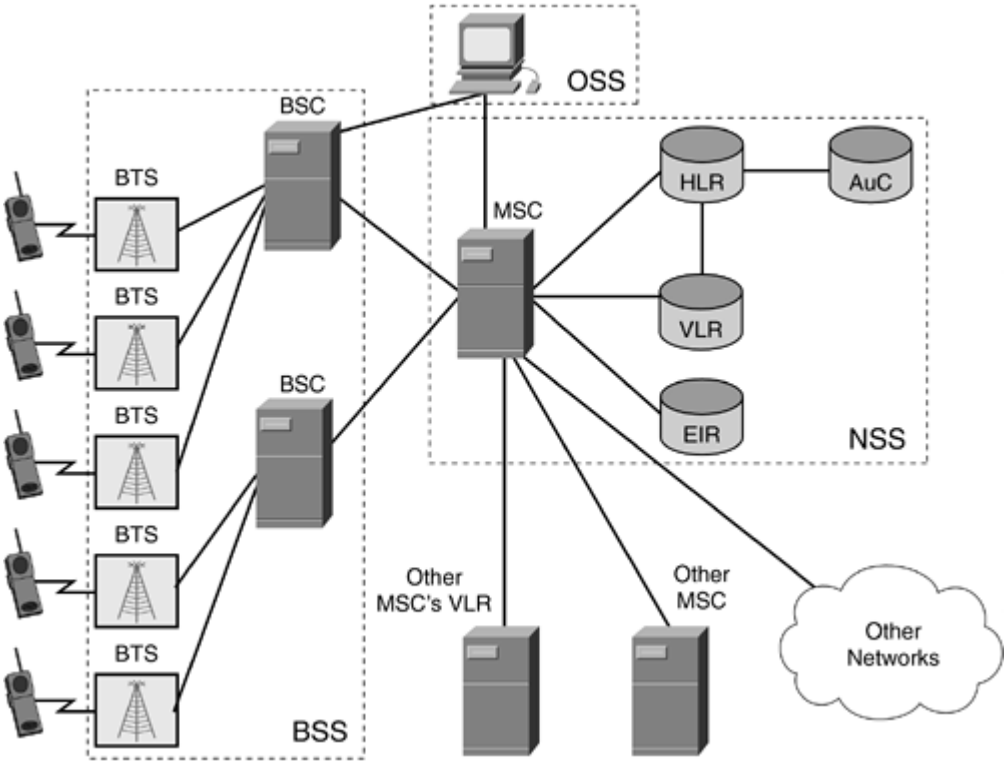
One benefit of 3G technology is that it unifies these diverse cellular standards. Although three different air interface modes exist—wideband CDMA, CDMA 2000, and the Universal Wireless Communication (UWC-136) interfaces—each should be able to work over both current GSM network architectures.

Network Architecture

GSM architecture can be divided into three broad functional areas: the Base Station Subsystem (BSS), the Network and Switching Subsystems (NSS), and the Operations Support Subsystem (OSS). Each of the subsystems is comprised of functional entities that communicate through various interfaces using specified protocols. The "Interfaces and Protocols" section of this chapter overviews the interfaces and SS7/C7 protocols that are used in the NSS and BSS.

Figure 12-1 shows a general GSM architecture to illustrate the scope and the entities that comprise the three subsystems.

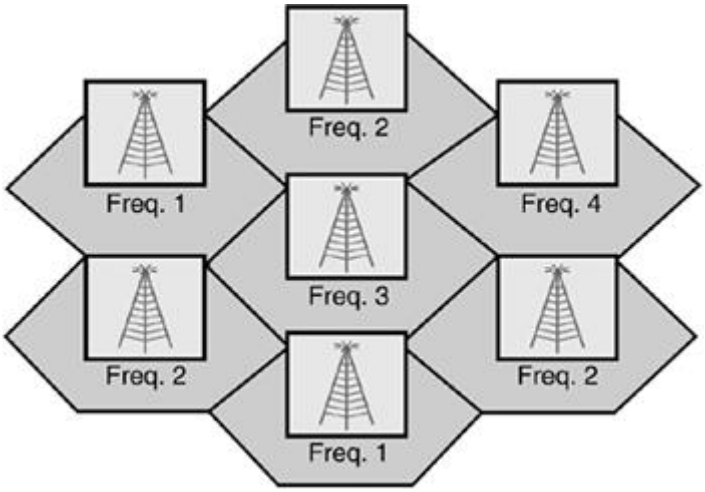
Figure 12-1. General GSM Architecture, Including the Three Main Separations in the Network



The BSS is comprised of the Base Transceiver Station (BTS) and the Base Station Controller (BSC). The BSS provides transmission paths between the Mobile Stations (MSs) and the NSS, and manages the transmission paths. The NSS is the brain of the entire GSM network and is comprised of the Mobile Switching Center (MSC) and four intelligent network nodes known as the Home Location Register (HLR), Visitor Location Register (VLR), Equipment Identity Register (EIR), and the Authentication Center (AuC). The OSS consists of Operation and Maintenance Centers (OMCs) that are used for remote and centralized operation, administration, and maintenance (OAM) tasks. The OSS provides means for a service provider to control and manage the network. The OSS is usually proprietary in nature and does not have standardized interfaces (using SS7 is irrelevant). Therefore, it is not considered. The BSS is the radio part, and this book does not detail radio related signaling. Therefore, the focus is on the NSS where the MAP protocol is used.

GSM utilizes a cellular structure. Each cell is hexagonal in shape so that the cells fit together tightly. Each cell is assigned a frequency range. The size of the cell is relatively small so the scarce frequencies can be reused in other cells. Each cell contains a base station, and a lot of planning goes into ensuring that base stations from different cells do not interfere with each other. One disadvantage of small cells is that the number of required base stations increases the infrastructure costs. The primary difference between GSM 900 and the GSM 1800/1900 systems is the air interface. In addition to using another frequency band, they both use a microcellular structure. As shown in Figure 12-2, this permits frequency reuse at closer distances, thereby enabling increases in subscriber density. The disadvantage is the higher attenuation of the air interface because of the higher frequency.

Figure 12-2. Frequency Reuse and Cellular Structure



One interesting point is that cell sizes vary because each cell can only serve a finite number of subscribers—typically 600 to 800. This means that cells become smaller for higher population density areas.

If a mobile moves from one cell to another during an active call, it should be clear that the call must be handed over to the new cell; this should be done in a fully transparent fashion to the subscriber. This process is known as a *handover*. The Mobile Switching Centre (MSC) monitors the strength of the incoming signal from the cellular phone (known as *MS*). When the signal power drops below a certain level, it indicates that the user might have entered another cell or is at the edge of the current cell. The MSC then checks to see if another cell is receiving a stronger cell. If it is, the call is transferred to that cell.

The approximate location of an MS, even if idle, has to be tracked to allow incoming calls to be delivered.

NOTE

Handovers and location tracking involve extensive and complex SS7/C7 signaling. In a cellular network, most signaling relates to the support of roaming functionality. Only a fraction of the signaling relates to call control.

The architecture that is presented in this section is not meant to be all-inclusive. Rather, its purpose is to provide the reader with the basic knowledge to comprehend SS7/C7 protocols that relate to cellular networks. When "GSM" is stated, it includes DCS, PCS, and GPRS networks. The rest of this section discusses the function of the components that comprise the NSS and BSS, along with the cellular phone itself and the identifiers associated with it.

Mobile Station (MS)

GSM refers to the cellular handsets as MS. PCMIA cards are also available for laptops to allow data transfer over the GSM network, without the need for a voice-centric handset. The MS consists of the physical equipment that the subscriber uses to access a PLMN and a removable smart card, known as the SIM, to identify the subscriber.

GSM was unique to use the SIM card to break the subscriber ID apart from the equipment ID. The SIM card is fully portable between *Mobile Equipment* (ME) units. This allows many features that we take for granted, such as being able to swap MS simply by swapping our SIM card over. All functionality continues seamlessly, including billing, and the telephone number remains the same.

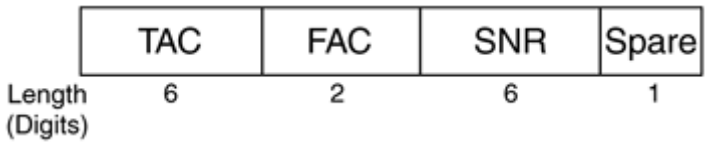
An MS has several associated identities, including the International Mobile Equipment Identity (IMEI), the International Mobile Subscriber Identity (IMSI), the Temporary Mobile Subscriber Identity (TMSI), and the Mobile Station ISDN (MSISDN) number. The following sections examine each of these identities, in turn, so that signaling sequences in which they are involved make sense.

IMEI

Each ME has a unique number, known as the IMEI, stored on it permanently. The IMEI is not only a serial number; it also indicates the manufacturer, the country in which it was produced, and the type approval. It is assigned at the factory.

GSM 03.03 specifies the IMEI, which is also defined by the 3GPP TS 23.003 [106]. The IMEI is used so actions can be taken against stolen equipment or to reject equipment that it cannot accept for technical and/or safety reasons. The IMEI allows tracing and prevention of fraudulent use and, in some circumstances, special network handling of specific MS types. Figure 12-3 shows the structure of the IMEI.

Figure 12-3. IMEI Structure



In the figure, the Type Approval Code (TAC) identifies the country in which the phone's type approval was sought, and its approval number. The first two digits of the TAC represent the country of approval. The Final Assembly Code (FAC) identifies the facility where the phone was assembled. Table 12-1 shows the codes that are currently in effect. The Serial Number (SNR) is an individual serial number that uniquely identifies each MS (within each TAC and FAC).

Table 12-1. Final Assembly Codes

Code	Facility
01, 02	AEG
07, 40	Motorola
10, 20	Nokia
30	Ericsson
40, 41, 44	Siemens
47	Option International
50	Bosch
51	Sony
51	Siemens
51	Ericsson
60	Alcatel
70	Sagem
75	Dancall
80	Philips
85	Panasonic

The IMEI is used for several fundamental network operations, such as when an MS is switched on; the IMEI number is transmitted and checked against a black/gray list. Operations that involve the IMEI are further discussed in later sections of this chapter.

In addition to current BCD coding, 3GPP is currently proposing to change the IMEI message structure to allow the use of hexadecimal coding. This would allow the production of 16.7 million mobile terminals with one TAC+FAC combination.

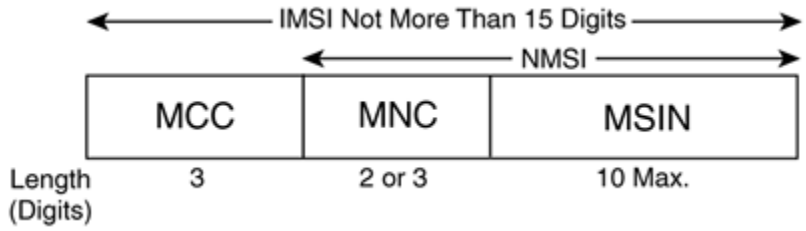
To display the IMEI on most MSs, enter ***#06#** on the keypad. This is useful for insurance purposes and allows the device to be blocked from network access, should it be stolen (network permitting).

IMSI

Each subscriber is assigned a unique number, which is known as the IMSI. The IMSI is the only absolute identity a subscriber has within GSM, and as such, it is stored on the SIM. The SIM is a credit size, or quarter-credit card size smart card that contains the subscriber's subscription details and grants the subscriber service when placed into a piece of ME. Among other purposes, it is used for subscriber billing, identification, and authentication when roaming.

The IMSI is specified in GSM 03.03, by 3GPP in TS 23.003, and the ITU in E.212. Figure 12-4 shows an IMSI's format.

Figure 12-4. IMEI Structure



In Figure 12-4, the Mobile Country Code (MCC) identifies the mobile subscriber's country of domicile. The Mobile Network Code (MNC) identifies the subscriber's home GSM PLMN.

The Mobile Station Identification Number (MSIN) identifies the mobile subscriber. The National Mobile Station Identity (NMSI) is the name given to MNC+MSIN fields.

The MCN's administration is the National Regulatory Authority's (NRAs) responsibility—for example, OFTEL in the UK or Telcordia in the USA—while network operators are usually responsible for the MSIN's arrangement and administration following the MNC assigned by the respective NRA. Appendix I contains a list of MCCs and MNCs.

TMSI

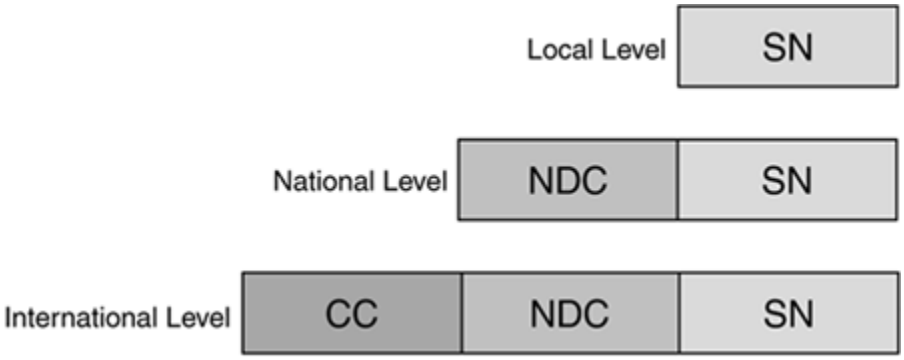
A TMSI is an alias used by the VLR (and the SGSN in GPRS enabled networks) to protect subscriber confidentiality. Please see section VLR for a description of the VLR. It is temporarily used as a substitute for the IMSI to limit the number of times the IMSI is broadcast over the air interface because intruders could use the IMSI to identify a GSM subscriber. TMSI is issued during the *location update* procedure. The VLR and SGSNs must be capable of correlating an allocated TMSI with the MS's IMSI to which it is allocated. The VLR assigns the TMSI to an MS during the subscriber's initial transaction with an MSC (for example, location updating). Because the TMSI has only local significance (within an area controlled by VLR), each network administrator can choose its structure to suit his needs. To avoid double allocation under failure/recovery conditions, it is generally considered good practice to make part of the TMSI related to time.

The TMSI is defined in 3GPP TS 23.003 [106].

MSISDN

MSISDN is the number the calling party dials to reach the called party—in other words, it is the mobile subscriber's directory number. This parameter refers to one of the ISDN numbers that is assigned to a mobile subscriber in accordance with ITU Recommendation E.213. A subscriber might have more than one MISDN on their SIM; examples include an MISDN for voice and an MISDN for fax. You can find additional MISDN details in GSM 03.02 and GSM 03.12. Figure 12-5 shows the format of an MSISDN.

Figure 12-5. MSISDN (E.164) Structure



In Figure 12-5, the National Destination Code (NDC) identifies the numbering area with a country and/or network/services. Country Code (CC) identifies a specific country, countries in an integrated NP, or a specific geographic area. Subscriber Number (SN) identifies a subscriber in a network or numbering area.

MSRN

The Mobile Station Roaming Number (MSRN) is solely used to route an incoming call. It is a temporary identifier that is used to route a call from the gateway MSC to the serving MSC/VLR.

The serving MSC/VLR is the MSC/VLR for the area where the subscriber currently roams. The VLR assigns an MSRN when it receives a request for routing information from the HLR. When the call has been cleared down, the MSRN is released back to the VLR.

Additional details about the MSRN can be found in GSM 03.03.

Subscriber Identity Module (SIM)

SIM cards are like credit cards and identify the user to the GSM network. They can be used with any GSM handset to provide phone access, ensure delivery of appropriate services to that user, and automatically bill the subscriber's network usage back to the home network.

As previously stated, GSM distinguishes between the subscriber and the MS. The SIM determines the subscriber's cellular number, thus permitting the subscriber to use other equipment (change MS) while maintaining one number and one bill. The SIM is a chip that is embedded in a card approximately the size of a credit card, or around a quarter of the size (the former tends to be outdated).

The SIM is the component that communicates directly with the VLR and indirectly with the HLR. These two critical networks components will be described later in this chapter.

Base Transceiver Station (BTS)

The base transceiver stations provide the connectively between the cellular network and the MS via the Airinterface. The BTS houses the radio transceivers that define a cell and handles the radio interface protocols with the mobile station.

Base Station Controller (BSC)

A number of BTSs are connected to the BSC on an interface that is known as the Abis interface.

It manages the radio interface channels, such as setup, release, frequency hopping, and handovers.

Mobile Switching Centre (MSC)

The MSC is the network subsystem's central component. Because a large number of BSCs are connected to an MSC, an MSC is effectively a regular ISDN switch that connects to the BSCs via the A-interface. The MSC provides routing of incoming and outgoing calls and assigns user channels on the A-interface.

It acts like a normal switching node of the PSTN or ISDN and provides all the necessary functionality for handling a mobile station, including registration, authentication, location updating, inter-MSC handovers, and call routing to a roaming subscriber.

The MSC also provides the connection to the public fixed networks.

Together with the MSC, the HLR and VLR provide GSM call routing and roaming capabilities.

Home Location Register (HLR)

The HLR can be regarded as a huge database that contains the information for hundreds of thousands of subscribers. Every PLMN has at least one HLR. While there is logically one HLR per GSM network, it might be implemented as a distributed database.

The HLR contains all administrative data that is related to each subscriber, who is registered in the corresponding GSM network, along with his current location. The location of each mobile station that belongs to the HLR is stored in order to be able to route calls to the mobile subscribers served by that HLR. The location information is simply the VLR address that currently serves the subscriber. An HLR does not have direct control of MSCs.

Two numbers that are attached to each mobile subscription and stored in the HLR include the IMSI and the MSISDN. The HLR also stores additional information, including the location information (VLR), supplementary services, basic service subscription information, and service restrictions (such as roaming permission). GSM 03.08 details the subscriber data's organization.

Visitor Location Register (VLR)

Like the HLR, the VLR contains subscriber data. However, it only contains a subset (selected administrative information) of the data that is necessary for call control and provision of the subscribed services for each mobile that is currently located in the geographical area controlled by the VLR. The VLR data is only temporarily stored while the subscriber is in the area that is served by a particular VLR. A VLR is responsible for one or several MSC areas. When a subscriber roams into a new MSC area, a location updating procedure is applied. When the subscriber roams out of the area that is served by the VLR, the HLR requests that it remove the subscriber-related data.

Although the VLR can be implemented as an independent unit, to date, all manufacturers of switching equipment implement the VLR with the MSC so the geographical area controlled by the MSC corresponds to that which is controlled by the VLR. The proximity of the VLR information to the MSC speeds up access to information that the MSC requires during a call.

Equipment Identity Register (EIR)

The EIR is a database that contains a list of all valid mobile equipment on the network. Each MS is identified by its IMEI. An IMEI is marked as invalid if it has been reported stolen or is not type approved.

The EIR contains a list of stolen MSs. Because the subscriber identity can simply be changed by inserting a new SIM, the theft of GSM MSs is attractive. The EIR allows a call bar to be placed on stolen MSs. This is possible because each MS has a unique IMEI.

Authentication Center (AuC)

The AuC is a protected database that stores a copy of the secret key that is stored in the subscriber's SIM card and is used for authentication and ciphering on the radio channel.

Serving GPRS Support Node (SGSN)

A SGSN is responsible for delivering data packets from and to the mobile stations within its geographical service area. Its tasks include packet routing and transfer, mobility management (attach/detach and location management), logical link management, and authentication and charging functions. The location register of the SGSN stores location information (such as current cell and current VLR) and user profiles (such as IMSI and address(es) used in the packet data network) of all GPRS users who are registered with this SGSN.

The SGSN delivers packets to mobile stations within its service area. SGSNs detect subscribers in their service area, query HLRs to obtain subscriber profiles, and maintain a record of their location.

Gateway GPRS Support Node (GGSN)

GGSNs maintain routing information that is necessary to tunnel the Protocol Data Units (PDUs) to the SGSNs that service specific mobile stations. Other functions include network and subscriber screening and address mapping.

Interfaces and Protocols

The previous section introduced GSM network architecture, and this section introduces the SS7/C7 protocols that are used. It also discusses interfaces, because different protocols are used on different interfaces. The SS7/C7 protocols MTP, SCCP, TUP, ISUP are protocols that were used before digital wireless networks were available. The final part of this section introduces SS7/C7 protocols that were specifically developed for GSM.

Table 12-2 summarizes the interfaces and protocols that are used in GSM.

Table 12-2. GSM Interfaces and Protocols

Interface	Between	Description
U_m	MS-BSS	The air interface is used for exchanges between a MS and a BSS. LAPDm, a modified version of the ISDN LAPD, is used for signaling.
Abis	BSC-BTS	This is a BSS internal interface that links the BSC and a BTS; it has not been standardized. The Abis interface allows control of radio equipment and radio frequency allocation in the BTS.
A	BSS-MSC	The A interface is between the BSS and the MSC. It manages the allocation of suitable radio resources to the MSs and mobility management. It uses the BSSAP protocols (BSSMAP and DTAP).
B	MSC-VLR	The B interface handles signaling between the MSC and the VLR. It uses the MAP/B protocol. Most MSCs are associated with a VLR, making the B interface "internal." Whenever the MSC needs to access data regarding an MS that is located in its area, it interrogates the VLR using the MAP/B protocol over the B interface.
C	GMSC-HLR or SMSG-HLR	The C interface is between the HLR and a GMSC or a SMSC. Each call that originates outside of GSM (such as an MS terminating call from the PSTN) must go through a gateway to obtain the routing information that is required to complete the call, and the MAP/C protocol over the C interface is used for this purpose. Also, the MSC can optionally forward billing information to the HLR after call clearing.
D	HLR-VLR	The D interface is between the HLR and VLR, and uses the MAP/D protocol to exchange data related to the location of the MS and

Table 12-2. GSM Interfaces and Protocols

Interface	Between	Description
		subsets of subscriber data.
E	MSC-MSC	The E interface connects MSCs. The E interface exchanges data that is related to handover between the anchor and relay MSCs using the MAP/E protocol. The E interface can also be used to connect the GMSC to an SMSC.
F	MSC-EIR	The F interface connects the MSC to the EIR and uses the MAP/F protocol to verify the status of the IMEI that the MSC has retrieved from the MS.
G	VLR-VLR	The G interface interconnects two VLRs of different MSCs and uses the MAP/G protocol to transfer subscriber information—for example, during a location update procedure.
H	MSC-SMSG	The H interface is located between the MSC and the SMSG and uses the MAP/H protocol to support the transfer of short messages. Again, GSM as well as ANSI-41 is unknown, but H in ANSI-41 is used for HLR–AC interface.
I	MSC-MS	The I interface is the interface between the MSC and the MS. Messages exchanged over the I interface are transparently relayed through the BSS.

In terms of the physical layer, the air interface (MS-BTS) uses RF radio transmission. The A-bis interface (BTS-BSC) uses 64 kbps over whatever medium is most convenient for installation: wire, optical, or microwave. All other interfaces in the GSM system use SS7/C7s MTP1 at the physical layer.

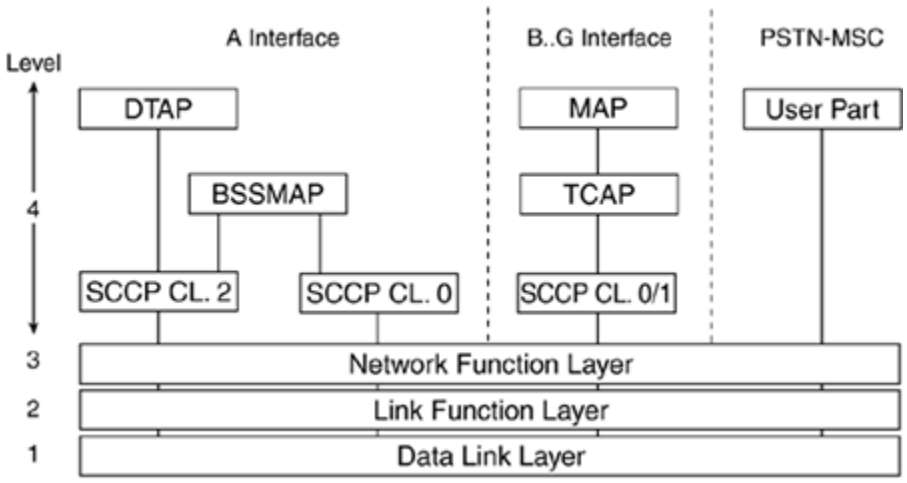
The data link layer that is used at the air interface (MS-BTS) is LAP-Dm; LAP-D is the data link layer that is used at the A-bis interface (BTS-BSC). All other interfaces in the GSM system use SS7/C7s MTP2 at the data link layer.

The air interface (MS-BTS) and the Abis interface (BTS-BSC) do not have a network layer. All other interfaces in the GSM system use SS7/C7s MTP3 and SCCP at the network layer.

The transport, session, and presentation layers are not used in SS7/C7—these functions are grouped together at the application layer, which is known as Level 4 in SS7/C7. GSM interfaces to fixed-line networks using ISUP or TUP (TUP is never used in North America).

Figure 12-6 shows the SS7 protocols that operate at each interface.

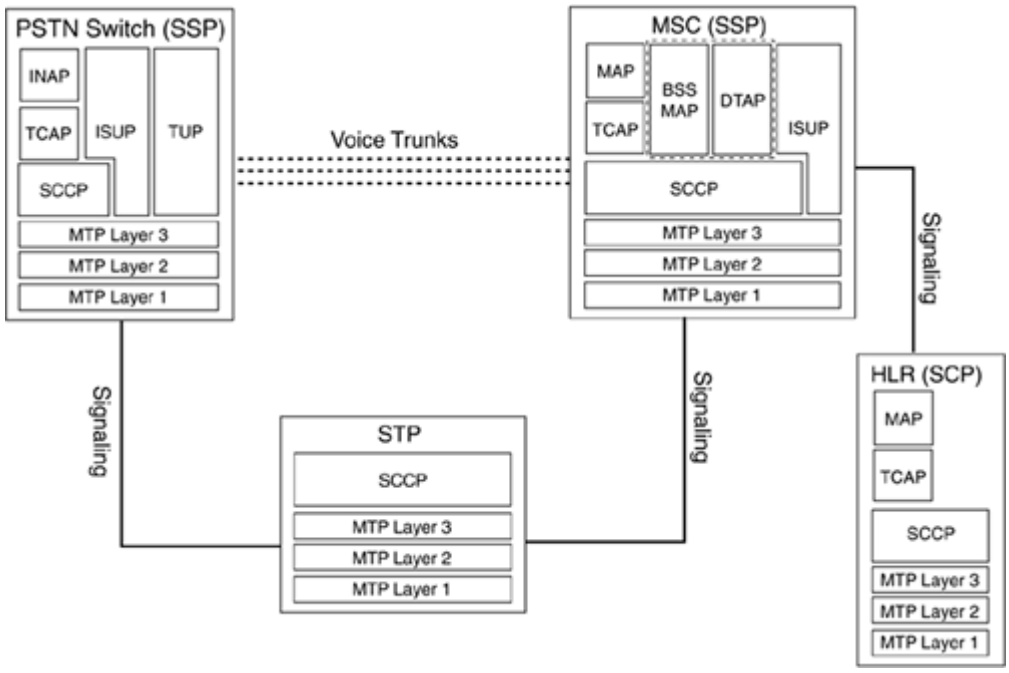
Figure 12-6. Protocols Operating at Each Interface



All of the interfaces around the MSC use SS7/C7-based protocols. The B, C, D, F, and G interfaces are referred to as MAP interfaces. These either connect the MSC to registers or connect registers to other registers. The E interface supports the MAP protocol and calls setup protocols (ISUP/ TUP). This interface connects one MSC to another MSC within the same network or to another network's MSC.

By this point, you can gather that different functional entities (e.g. HLR, MSC, and so on) run the required and therefore differing stack of SS7/C7 protocols. In relation to the following diagram, remember that the MSC runs MAP-MSC, and that MAP-VLR and the HLR run MAP-HLR.

Figure 12-7. Protocols Required for Functional Entities



BSSAP (DTAP/BSSMAP)

On the A interface, an application part known as the BSSAP is used. BSSAP can be further separated into the base station subsystem management application part (BSSMAP) and the direct transfer application part (DTAP).

Neither the BTS nor the BSC interpret CM and MM messages. They are simply exchanged with the MSC or the MS using the DTAP protocol on the A interface. RR messages are sent between the BSC and MSC using the BSSAP.

BSSAP includes all messages exchanged between the BSC and the MSC that the BSC actually processes—examples include PAGING, HND_CMD, and the RESET message. More generally, BSSAP comprises all messages that are exchanged as RR messages between MSC and BSC, and messages that are used for call-control tasks between the BSC and the MSC.

The DTAP comprises all messages that the subsystem of the NSS and the MS exchange. DTAP transports messages between the MS and the MSC, in which the BSC has just the relaying function.

Mobile Application Part (MAP)

The MAP is an extension of the SS7/C7 protocols that are added to support cellular networks. It defines the operations between the MSC, the HLR, the VLR, the EIR, and the fixed-line network. It comes in two incompatible variants: GSM-MAP and ANSI-41 MAP. While GSM-MAP only supports GSM, ANSI-41 supports AMPS, NAMPS, D-AMPS/TDMA, CDMA

(cdma One and cdma 2000), and GSM. GSM-MAP is the international version, while ANSI-41 is the North American version.

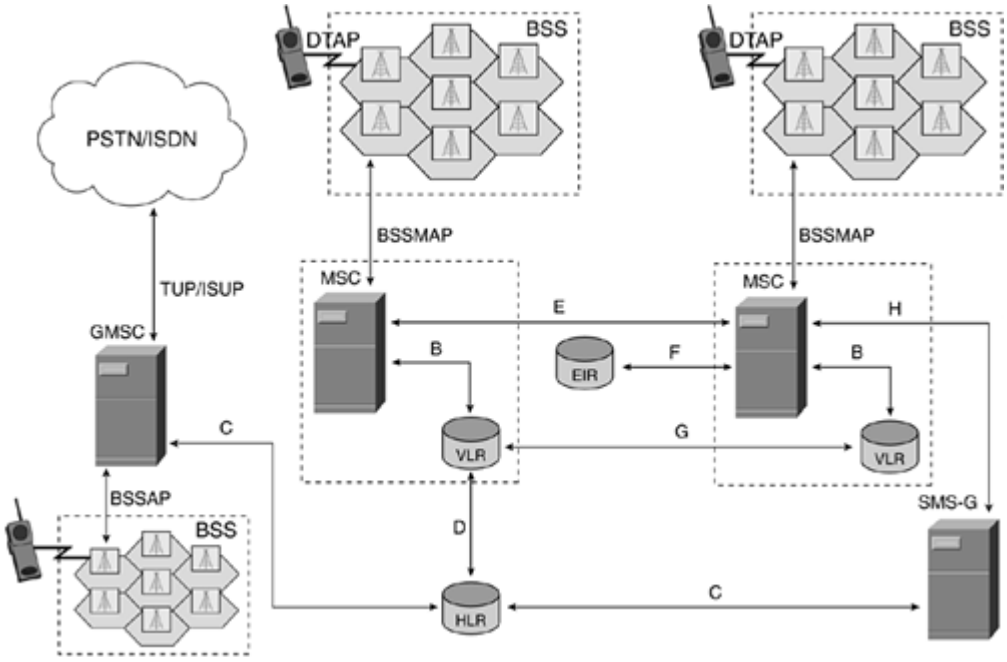
The MAP is used to define the operations between the network components (such as MSC, BTS, BSC, HLR, VLR, EIR, MS, and SGSN/GGSN in GPRS). This involves the transfer of information between the components using noncircuit-related signaling. MAP signaling enables location updating, handover, roaming functionality, authentication, incoming call routing, and SMS. MAP specifies a set of services and the information flows between GSM components to implement these services. MAP can be considered an extension of the SS7/C7 protocol suite created specifically for GSM and ANSI-41 networks.

MAP uses TCAP over SCCP and MTP. TCAP correlates between individual operations. The TCAP transaction sublayer manages transactions on an end-to-end basis. The TCAP component sublayer correlates commands and responses within a dialog. Chapter 10, "Transaction Capabilities Application Part (TCAP)," describes TCAP in more detail.

MAP protocols are designated MAP/B–MAP/H, according to the interface on which the protocol functions. For example, the MAP signaling between the GMSC and the HLR is MAP/F.

Figure 12-8 shows the specific MAP-n protocols. The PCS 1900 specifications use the same MAP interfaces, but PCS 1900 also defines MAP-H.

Figure 12-8. MAP-n Protocols



MAP allows implementation of functions such as location updating/roaming, SMS delivery, handover, authentication, and incoming call routing information. The MAP protocol uses the TCAP protocol to transfer real-time information (between NSS components).

- MAP provides the functionality to route calls to and from the mobile subscribers—it has the mechanisms necessary for transferring information relating to subscribers roaming between network entities in the PLMN.
- The U.S. version is known as ANSI-41-MAP (standardized by EIA/TIA).
- The international version is known as GSM-MAP (standardized by ITU/ETSI).

MAP only makes use of the connectionless classes (0 or 1) of the SCCP.

Table 12-4 shows the SCCP Subsystem Numbers (SSNs) that are specified for MAP.

Table 12-3. SSNs Used by MAP

SCCP Subsystem Numbers	Use
0 0 0 0 0 1 0 1	For the entire MAP (reserved for possible future use)
0 0 0 0 0 1 1 0	HLR
0 0 0 0 0 1 1 1	VLR
0 0 0 0 1 0 0 0	MSC
0 0 0 0 1 0 0 1	EIR
0 0 0 0 1 0 1 0	Allocated for evolution (possible Authentication centre)

Mobility Management and Call Processing

This section provides an introductory overview of mobility management (i.e., allowing a subscriber to roam) and call processing (the setting up and clearing down of calls) in GSM networks.

Mobility management entails keeping track of the MS while it is on the move. The mobility management procedures vary across three distinct scenarios, namely:

- MS is turned off
- MS is turned on but is idle
- MS has an active call

In the first scenario, when it cannot be reached by the network because it does not respond to the paging message, the MS is considered to be in the turned-off state. In this scenario, the MS obviously fails to provide any updates in relation to changes in Location Area (LA), if any exist. In this state, the MS is considered detached from the system (IMSI detached).

In the second scenario, the MS is in the ready state to make or receive calls. The system considers it attached (IMSI attached), and it can be successfully paged. While on the move,

the MS must inform the system about any changes in LA; this is known as location updating.

In the third scenario, the system has active radio channels that are allowed to the MS for conversation/data flow. The MS is required to change to new radio channels if the quality of current channels drops below a certain level; this is known as handover. The MSC (sometimes BSC) makes the decision to handover an analysis of information that is obtained real-time from the MS and BTS.

All operations revolve around the three scenarios presented above. The rest of this chapter examines these operations in more detail, beginning with simple operations: paging, IMSI detach/attach. Following, more complex operations are presented, such as location update, call handover, mobile terminated call, mobile originated call, and mobile-to-mobile call.

Location Update

Location updating is the mechanism that is used to determine the location of an MS in the idle state. The MS initiates location updating, which can occur when:

- The MS is first switched on
- The MS moves within the same VLR area, but to a new LA
- The MS moves to a new VLR area
- A location updated timer expires

Mobile Terminated Call (MTC)

In the case of an MTC, a subscriber from within the PSTN dials the mobile subscriber's MSISDN. This generates an ISUP IAM message (it also could potentially be TUP as Level 4) that contains the MSISDN as the called party number. The ISDN (i.e., PSTN) routes the call to the GMSC in the PLMN, based on the information contained in the MSISDN (national destination code and the country code).

The GMSC then identifies the subscriber's HLR based upon the MSISDN and invokes the MAP/C operation *Send Routing Information (SRI)* towards the HLR to locate the MS. The SRI contains the MSISDN. The HLR uses the MSISDN to obtain the IMSI.

Because of past location updates, the HLR already knows the VLR that currently serves the subscriber. The HLR queries the VLR using the MAP/D operation Provide Roaming Number (PRN) to obtain the MSRN. The PRN contains the subscriber's IMSI.

The VLR assigns a temporary number known as the *mobile station roaming number* (MSRN), which is selected from a pool, and sends the MSRN back in an MAP/D *MSRN Acknowledgement* to the HLR.

The HLR then passes the MSRN back to the GMSC in a MAP/C Routing Information Acknowledgement message. To the PSTN, the MSRN appears as a dialable number.

Since the GMSC now knows the MSC in which the MS is currently located, it generates an IAM with the MSRN as the called party number. When the MSC receives the IAM, it

recognizes the MSRN and knows the IMSI for which the MSRN was allocated. The MSC then returns the MSRN to the pool for future use on another call.

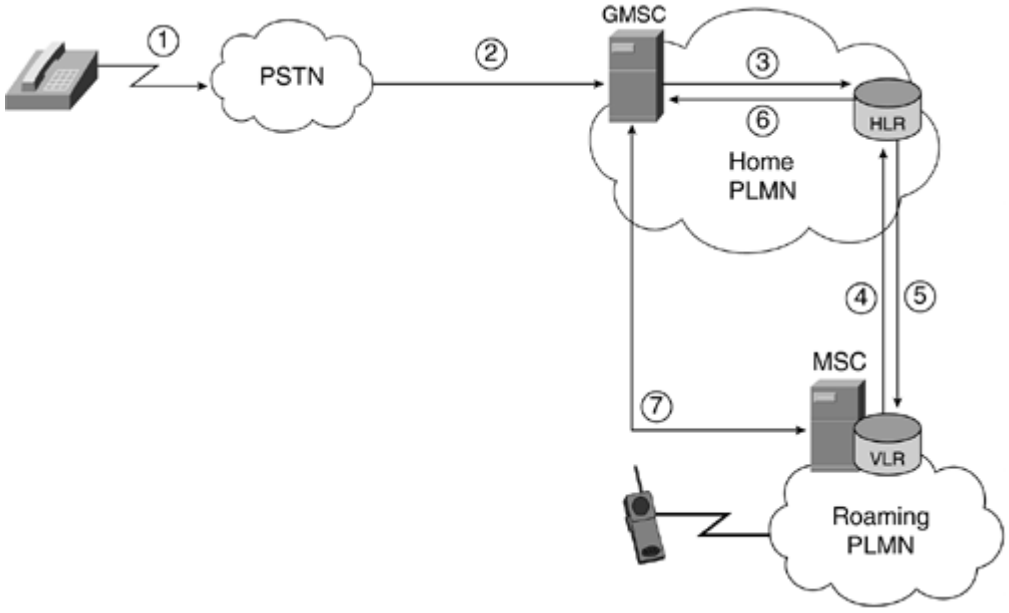
The MSC sends the VLR a MAP/B Send Information message requesting information, including the called MS's capabilities, services subscribed to, and so on. If the called MS is authorized and capable of taking the call, the VLR sends a MAP/B Complete Call message back to the MSC.

The MSC uses the LAI and TMSI received in the Complete Call message to route a BSSMAP Page message to all BSS cells in the LA.

Air interface signaling is outside the scope of this book.

Figure 12-9 shows the sequence of events involved in placing an MTC.

Figure 12-9. Placing an MTC



In Figure 12-9, the sequence of events involved in placing an MTC is as follows:

1. The calling subscriber uses the MSISDN to dial the mobile subscriber.
2. The MSISDN causes the call to be routed to the mobile network gateway MSC (GMSC).
3. The GMSC uses information in the called number digits to locate the mobile subscriber's HLR.
4. The HLR has already been informed about the location (VLR address) for the mobile subscriber; it requests a temporary routing number to allow the call to be routed to the correct MSC.

5. The MSC/VLR responds with a temporary routing number that is only valid for the duration of this call.
6. The routing number is returned to the GMSC.
7. The call is made using ISUP (or TUP) signaling between the GMSC and the visited MSC.

If the calling subscriber were in the same PLMN as the called party (internal MS-to-MS call), steps 2 and 3 would not be required.

Chapter 13 describes GSM-MAP operations in more detail. Appendix F, "GSM and ANSI MAP Operations," provides a list of GSM-MAP operations.

Summary

Cellular networks have undergone a rapid development phase since their initial introduction in the early 1980s. Modern cellular networks are digital and use SS7 for communication between network entities. GSM is the most popular digital cellular standard. GSM management call control, subscriber mobility, and text messaging (SMS) use a SS7 subsystem known as MAP. MAP provides operations for tracking the subscriber's location to deliver a call, signal the subscriber's intention to place a call, and deliver text messages between handsets. Operations and maintenance staff also use it to change the subscriber's profile—to add or revoke services.

Chapter 13. GSM and ANSI-41 Mobile Application Part (MAP)

In fixed-line networks, the subscriber's location is static and specified according to the numbering scheme used in the network.

In cellular telephony systems, the subscriber's location can change drastically without the system being aware—for example, the subscriber might switch his cell phone off just before boarding a plane, and then switch it back on in a new country. For incoming calls to mobile subscribers, there is no direct relationship between the subscriber's location and the cell phone number. Because the location and other information must be derived real-time before a call can be delivered to a cell phone, such mobile terminating calls require the performance of a large amount of initial noncircuit-related signaling.

In contrast, mobile-originated calls (outgoing calls) place far less initial signaling overhead because the radio system to which the subscriber is connected knows the subscriber's location. Furthermore, because a subscriber is on the move, the base transceiver system (BTS), the base station controller (BSC), and even the mobile switching centre (MSC) can change. These changes require a lot of noncircuit-related signaling, particularly if the subscriber is currently engaged in a call—the subscriber should not be aware that such handovers between cellular network equipment takes place.

Retrieving the subscriber's profile is also a straightforward task for fixed-line networks because it resides at the subscriber's local exchange. In cellular networks, the ultimate exchange (MSC) to which the mobile subscriber is connected changes because the

subscriber is mobile, and it would be completely unmanageable to place the subscriber's profile (which might change) at every MSC throughout the world.

It is primarily for these reasons that cellular networks contain two databases, known as the *Home Location Register (HLR)* and the *Visitor Location Register (VLR)*, in addition to the cellular-specific switch known as the MSC. For a description of the nodes used in a Global System for Mobile communications (GSM) network, see Chapter 12, "Cellular Networks."

Mobile application part (*MAP*) is the protocol that is used to allow the GSM network nodes within the Network Switching Subsystem (NSS) to communicate with each other to provide services, such as roaming capability, text messaging (SMS), and subscriber authentication. MAP provides an application layer on which to build the services that support a GSM network. This application layer provides a standardized set of operations. MAP is transported and encapsulated with the SS7 protocols MTP, SCCP, and TCAP.

This chapter specifies the MAP operations (or messages) that are used in GSM Phase 2. A small number of operations have been added to support General Packet Radio Service (GPRS) and 3rd Generation (3G) Universal Mobile Telecommunications System (UMTS), but they are beyond the scope of this book.

See Appendix F, "GSM and ANSI MAP Operations," for a list of the MAP operations used in GSM.

MAP Operations

MAP Phase 2 operations can be divided into the following main categories, which are addressed in this chapter:

- Mobility Management
- Operation and Maintenance
- Call Handling
- Supplementary Services
- Short Message Service

The chapter ends with a summary of GSM and ANSI MAP operations.

Mobility Management

Mobility management operations can be divided into the following categories:

- Location Management
- Paging and Search
- Access Management
- Handover
- Authentication Management
- Security Management
- IMEI Management
- Subscriber Management
- Identity Management
- Fault Recovery

The following section examines the MAP operations that are used in each of these categories, excluding Paging and Search, Access Management, Security Management and Identity Management because these categories were removed at Phase 2.

Location Management

To minimize transactions with the HLR, it only contains location information about the MSC/VLR to which the subscriber is attached. The VLR contains more detailed location information, such as the location area in which the subscriber is actually roaming. SeeChapter 12, "Cellular Networks," for more information about location areas. As a result, the VLR requires that its location information be updated each time the subscriber changes location area. The HLR only requires its location information to be updated if the subscriber changes VLR.

Location management operations include the following:

- updateLocation
- cancelLocation
- sendIdentification
- purgeMS

updateLocation

This message is used to inform the HLR when an MS (in the idle state) has successfully performed a location update in a new VLR area. In this way, the HLR maintains the location of the MS (VLR area only). In Appendix L, "Tektronix Supporting Traffic," Figure 13-3 contains a trace that shows an HLR's decode calling a VLR (to perform cancel location). In Figure 13-1, the MS has roamed from a VLR area that is controlled by VLR-A to an area that is controlled by VLR-B. Note that the purgeMS operation is optional in a location update procedure.

Figure 13-3. MAP Operation Sequences in a Handover

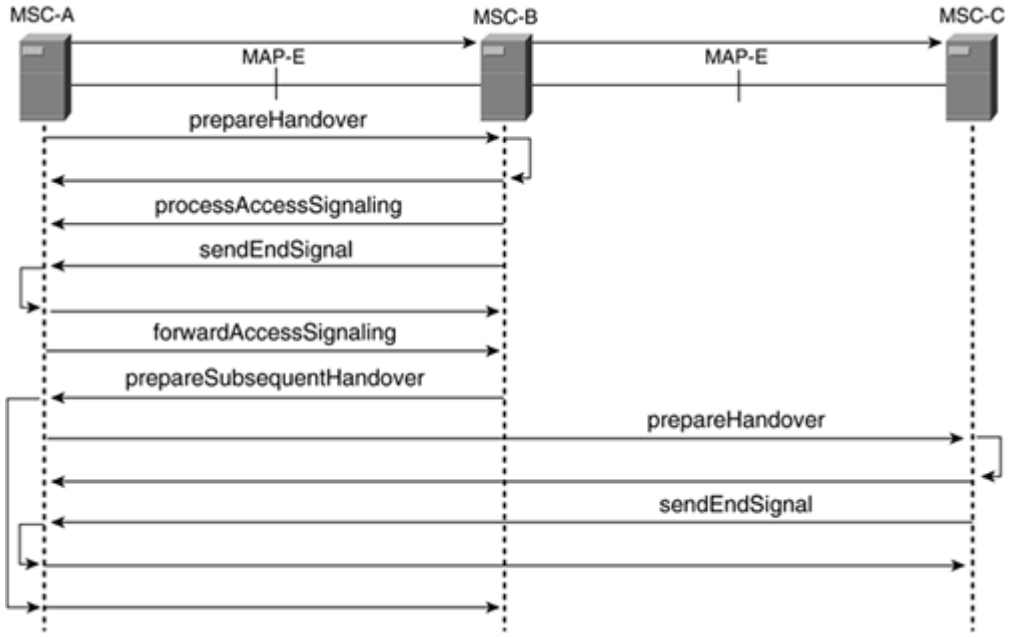
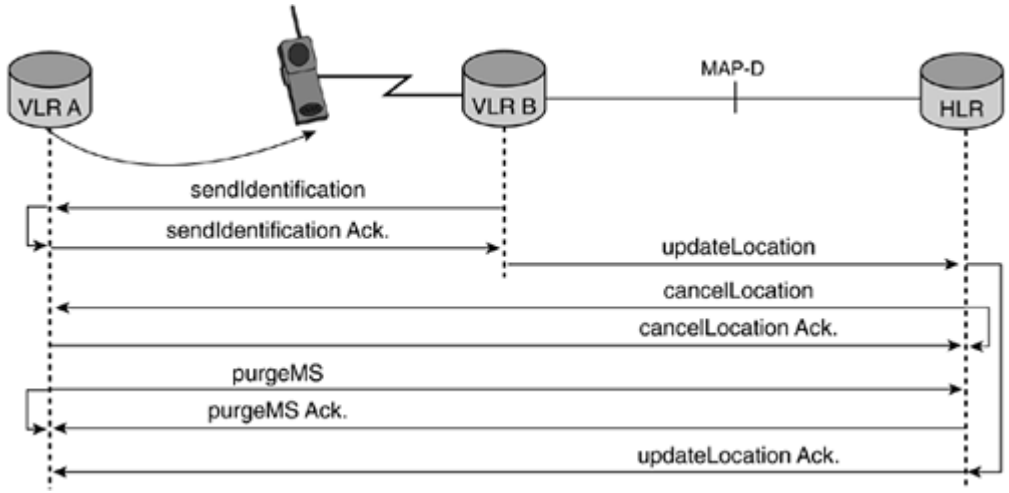


Figure 13-1. Showing the MAP Operation Sequences Involved in a Location Update



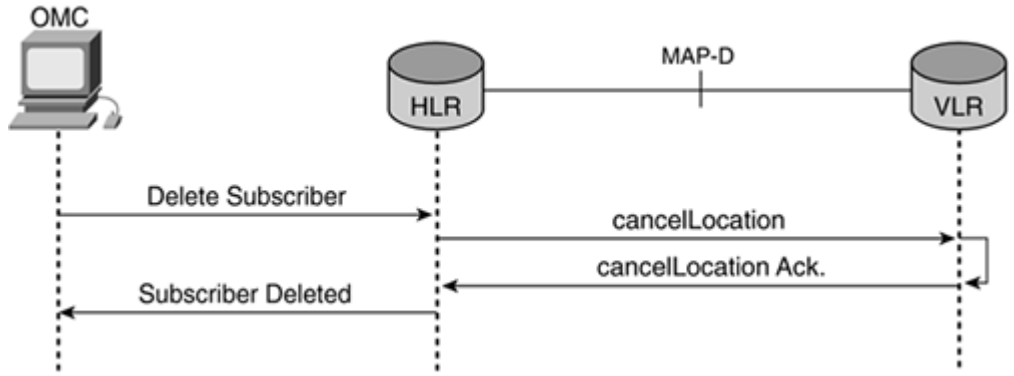
cancelLocation

The cancelLocation operation is used to delete a subscriber's profile from the previous VLR, following registration with a new VLR—in other words, following an updateLocation. When the HLR receives an updateLocation from a VLR other than the one that is currently stored in its tables, it sends a cancelLocation to the old VLR. The cancelLocation includes the International Mobile Subscriber Identity (IMSI) and the Local Mobile Subscriber Identity (LMSI) to identify the subscriber whose profile should be deleted as parameters. For details

of the IMSI and LMSI see Chapter 12, "Cellular Networks." In Appendix L, "Tektronix Supporting Traffic," Example L-3 contains a trace that shows an HLR's decode calling a VLR (to perform cancel location).

Operators can also use the operation to impose roaming restrictions following a change in the subscriber's subscription. It is also used as part of the process of completely canceling a subscriber's subscription. When the HLR receives a request from the Operation and Maintenance Center (OMC) to delete the subscriber, the HLR deletes the subscriber's data and sends a cancelLocation to the VLR that serves the subscriber. Figure 13-2 shows a subscriber's subscription being cancelled, thereby disabling their service.

Figure 13-2. MAP Operation Sequences in Which a Subscriber's Service is Disabled



In addition, a cancelLocation operation is sent from the HLR to the VLR if the authentication algorithm or authentication key of the subscriber is modified.

sendIdentification

When the MS changes to a new VLR area, the new VLR queries the old VLR using a sendIdentification operation to obtain authentication information. The sendIdentification operation sends the TMSI as its argument, and the result contains the IMSI and other authentication information (RAND, SRES, and optionally KC). If it is unable to obtain this information, it can retrieve the information from the HLR via a sendAuthenticationInfo operation.

purgeMS

This message is sent if an MS has been inactive (no call or location update performed) for an extended period of time. The VLR sends this message to the HLR to indicate that it has deleted its data for that particular MS. The HLR should set a flag to indicate that the MS should be treated as not reached; as a result, the HLR no longer attempts to reach the MS in the case of a mobile terminated call or a mobile terminated short message.

Handover

Handover between MSCs is known as inter-MSC handover: basic inter-MSC handover and subsequent inter-MSC handover. A basic inter-MSC handover is where the call is handed from the controlling MSC (MSC-A) to another MSC (MSC-B). A subsequent inter-MSC handover is an additional inter-MSC handover during a call. After a call has been handed over from MSC-A to MSC-B, another handover takes place, either to a new MSC (MSC-C) or back to the original MSC (MSC-A).

The following sections describe these MAP handover operations:

- prepareHandover
- sendEndSignal
- processAccessSignalling
- forwardAccessSignalling
- prepareSubsequentHandover

prepareHandover

The prepareHandover message is used to carry a request and response between the two MSCs at the start of a basic inter-MSC handover (MSC-A to MSC-B). It is used to exchange BSSAP messages, such as HAN_REQ and HAN_ACK, for this purpose. It is the decision of MSC-A to hand over to another MSC. The prepareHandover message does not contain subscriber information—only information that is necessary for MSC-B to allocate the necessary radio resources and possibly some optional information, such as an IMSI.

sendEndSignal

Following a successful inter-MSC handover (from MSC-A to MSC-B in the case of a basic handover), MSC-B sends a sendEndSignal message to MSC-A to allow it to release its radio resources. If the call was originally established with MSC-A, it keeps control of the call and is known as the *anchor* MSC following the handover. As a result, MSC-B does not receive information about the release of the call. To solve this problem, MSC-A sends a sendEndSignal to MSC-B to inform it that it can release its own radio resources.

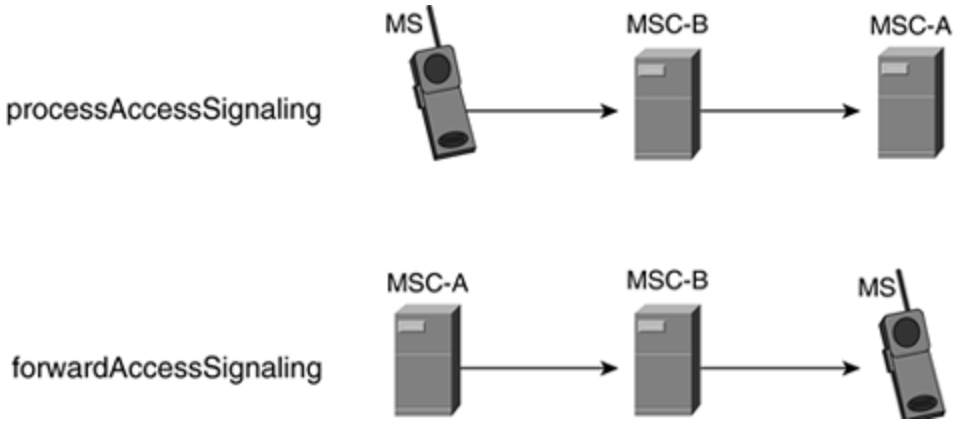
processAccessSignaling

The messages processAccessSignaling and forwardAccessSignaling are used to pass BSSAP messages between the MS and the anchor MSC transparently and between the anchor MSC and the MS, respectively. As stated previously, MSC-A keeps control of the call after a successful inter-MSC handover from MSC-A to MSC-B. The BSSAP messages travel from the MS to MSC-A via MSC-B. The message processAccessSignaling carries data from the MS to MSC-A and is sent from MSC-B to MSC-A. The message forwardAccessSignaling is the reverse; it carries data from MSC-A to the MS via MSC-B, as shown in Figure 13-3.

forwardAccessSignaling

See processAccessSignaling. If call control information is required to be passed to the serving MSC (MSC-B), the anchor (controlling MSC, MSC-A) sends the information using a forwardAccessSignaling message.

Figure 13-4. Direction of processAccessSignaling and forwardAccessSignaling



prepareSubsequentHandover

If another inter-MSC is required (back to MSC-A or to another MSC, C), then MSC-B sends this message to MSC-A. It contains the information required for MSC-A to send a prepareHandover message to MSC-C. Refer to Figure 13-3.

Authentication Management

MAP operation sendIdentificationInfo is the only operation in Phase 2 that falls under the category of authentication management. See sendIdentification for a description of this operation.

IMEI Management

The only MAP operation in the IMEIs management category is checkIMEI, which is used to check whether a piece of mobile equipment is on a black, gray, or white list. To perform an IMEI check, the serving MSC requests that the MS provide its IMEI. On receiving the IMEI from the MS, the MSC sends the IMEI to the EIR in a MAP checkIMEI operation. The EIR checks the status of the IMEI and sends the result back to the MSC. The equipment status can be white listed, gray listed, blacklisted, or unknown.

Blacklisted equipment is equipment that has been reported stolen and is, therefore, not granted permission to use the network (barred). If the status indicates that the equipment is blacklisted, an alarm might be generated on the operation and maintenance interface; this is network operator-dependent. The network operator can use the gray listed equipment list to block a certain model of equipment (or even a particular software version) from using his network if, for example, a certain handset type has proven to act erroneously on the network. Gray listed equipment cannot be barred; instead, it can be chosen to track

the equipment for observation purposes. The white list contains all the equipment identities that are permitted for use and to which service should therefore be granted.

Criminals have been able to change mobile handsets' IMEI fairly easily using a data cable (to connect it to a PC) and specialist software. Because of this and the abundance and the high price of mobile handsets, theft has hit epidemic levels in many parts of the world. Recently, the United Kingdom passed legislation known as the Mobile Telephones (Re-programming) Act making it illegal to reprogram the IMEI, and manufacturers were pressed (with limited success) to make the IMEI tamper-proof. In addition, the operators and the GSM association set up a nationwide EIR, known simply as the Central Equipment Identity Register (CEIR) so that stolen mobile equipment could be reported as easily as a stolen credit card. Before CEIR, if the equipment had been blacklisted with one operator, in most cases you could simply put in an SIM card for another operator because the operators failed to pool information.

Subscriber Management

An HLR uses subscriber management procedures to update a VLR with specific subscriber data when the subscriber's profile is modified. A subscriber's profile can be modified, because the operator has changed the subscription of the subscriber's basic services or one or more supplementary services. A subscriber's profile might also be modified, because the subscriber himself has activated or deactivated one or more supplementary services.

Subscriber management uses the insertSubscriberData and deleteSubscriberData operations.

insertSubscriberData

The HLR uses the insertSubscriberData operation to provide the VLR with the current subscriber profile—for example, during a location update or restore data procedure. It is also used if the operator (via the OMC) or the subscriber himself modifies the data—for example, barring all or certain types of calls. The operation insertSubscriberData is sent as many times as necessary to transfer the subscriber data from the HLR to the VLR.

deleteSubscriberData

The HLR uses the deleteSubscriberData operation to inform the VLR that a service has been removed from the subscriber profile. The subscriber might have subscribed to a number of services, such as international roaming. The operator can use this operation to revoke such subscriptions.

Fault Recovery

The fault recovery procedures ensure that the subscriber data in the VLR becomes consistent with the subscriber data that is stored in the HLR for a particular MS, and that the MS location information in the HLR and VLR is accurate following a location register fault.

3GPP TS 23.007 gives the detailed specification of fault recovery procedures of location registers.

The fault recovery procedures use the following three MAP operations:

- reset
- forwardCheckSsIndication
- restoreData

reset

The HLR that returns to service following an outage sends this operation to all VLRs in which that HLR's MSs are registered according to any available data following the outage.

forwardCheckSsIndication

This operation is optionally sent to all MSs following an HLR outage. The MSs are requested to synchronize their supplementary service data with that which is held in the HLR.

restoreData

When a VLR receives a provideRoamingNumber request from the HLR for either an IMSI that is unknown to the VLR or an IMSI in which the VLR entry is unreliable because of an HLR outage, the VLR sends a restoreData message to the HLR to synchronize the data.

Operation and Maintenance

Operation and maintenance can be divided into the following categories:

- Subscriber Tracing
- Miscellaneous

The following sections review the MAP operations that are used in each of these categories.

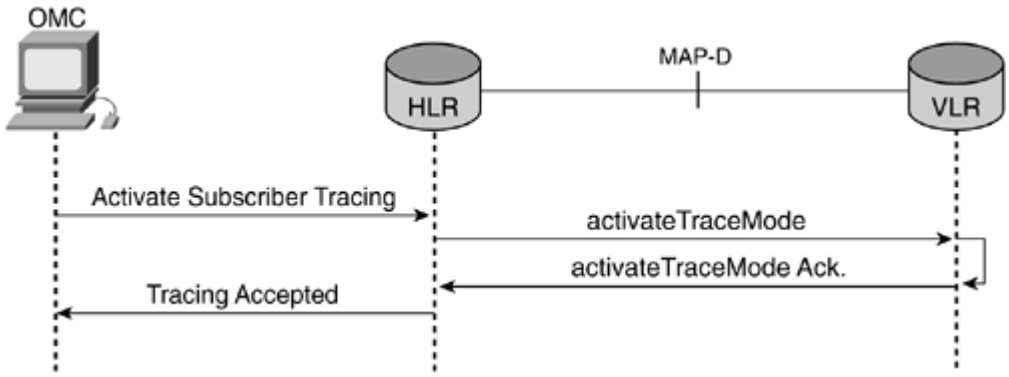
Subscriber Tracing

Subscriber tracing has two operations: activateTraceMode and deactivateTraceMode.

activateTraceMode

The HLR uses activateTraceMode to activate trace (subscriber tracking) mode for a particular subscriber (IMSI); the OSS requests activateTraceMode. The VLR waits for that particular MS to become active, at which time it sends a request to its MSC to trace the MS.

Figure 13-5. MAP Operation Sequence to Initiate and Terminate Subscriber Tracing



deactivateTraceMode

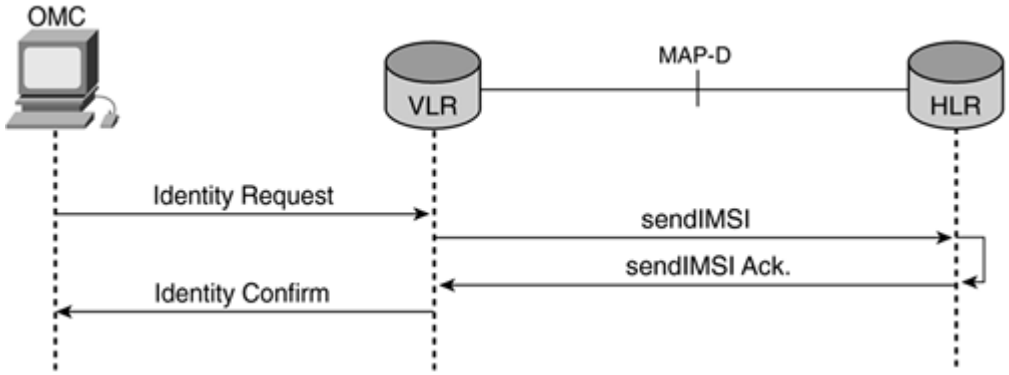
Upon receiving this message, the HLR turns off the trace mode and sends the message to the VLR, which also disables trace mode for that particular subscriber. See activateTraceMode.

Miscellaneous

The only operation in the Miscellaneous subcategory is sendIMSI.

Following the OMC's request to the VLR to identify a subscriber based on his Mobile Subscriber ISDN Number (MSISDN), the VLR and HLR exchange sendIMSI messages. If the MSISDN cannot be identified, an unknown subscriber indication is passed to the VLR. Otherwise, the IMSI is obtained from the HLR and returned to the VLR.

Figure 13-6. MAP Operation Sequence When an Operations and Management Center (OMC) Requests Subscriber Identity



Call Handling

The call handling procedures primarily retrieve routing information to allow mobile terminating calls to succeed. When a mobile originating or a mobile terminating call has reached the destination MSC, no further MAP procedures are required.

Other procedures performed by MAP's call handling routines include the restoration of call control to the Gateway Mobile Switching Center (GMSC) if the call is to be forwarded. In addition, the call handling routing processes the notification that the remote user is free for the supplementary service message call completion to busy subscribers (CCBS).

Call handling does not have subcategories of operations; it simply has the following two operations:

- sendRoutingInfo
- provideRoamingNumber

In the case of an MTC, a subscriber from within the PSTN/ISDN dials the mobile subscriber's MSISDN, thereby generating an ISUP IAM message (alternatively, TUP could be used) that contains the MSISDN as the called party number. Based on the information contained in the MSISDN (national destination code and the country code), the PSTN/ISDN routes the call to the GMSC in the PLMN.

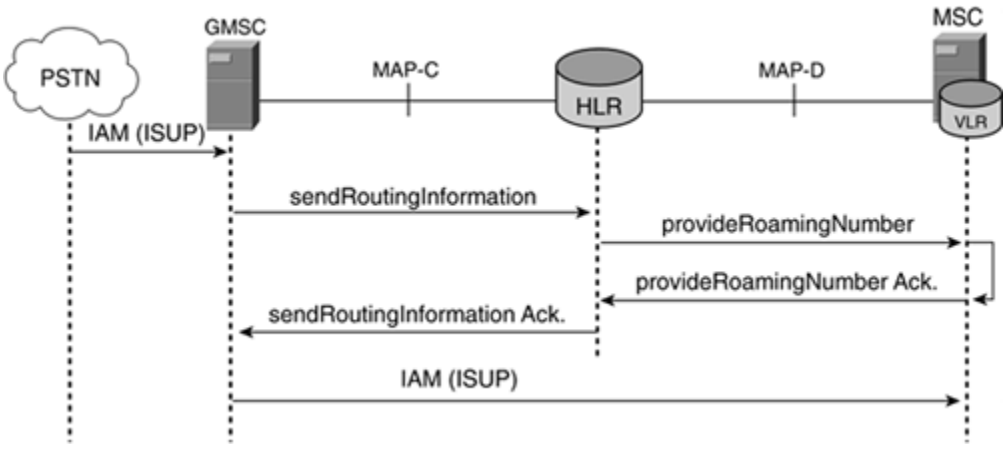
The GMSC then identifies the subscriber's HLR based on the MSISDN, and invokes the MAP operation sendRoutingInformation with the MSISDN as a parameter towards the HLR to find out where the MS is presently located.

Because of past location updates, the HLR already knows the VLR that currently serves the subscriber. To obtain a mobile station roaming number (MSRN), the HLR queries the VLR using the operation provideRoamingNumber with the IMSI as a parameter. The VLR assigns an MSRN from a pool of available numbers and sends the MSRN back to the HLR in an acknowledgement.

Because the GMSC now knows the MSC in which the MS is currently located, it generates an IAM with the MSRN as the called party number. When the MSC receives the IAM, it recognizes the MSRN and knows the IMSI for which the MSRN was allocated. The MSRN is then returned to the pool for use on a future call.

Figure 13-7 shows how the routing information is obtained to route the call from the calling parties exchange to the called parties exchange (serving MSC).

Figure 13-7. MAP Operations When the GMSC Requests a Routing Number for the MSC When the Subscriber is Roaming



The BSSAP PAGE message is used for contacting all BSS cells in the location area (LA) when searching for the MS. The radio-related signaling is outside the scope of this book; however, this book does reference radio-related messages that are required for understanding NSS signaling. When the MS responds with a DTAP ALERT message, the serving MSC sends an ISUP ACM back to the GMSC, which forwards it to the calling subscriber's PSTN/ISDN switch. When the called subscriber accepts the call, the MS sends a DTAP CON message to the serving MSC that, in turn, sends an ISUP ANM message back to the calling party's PSTN/ISDN switch through the GMSC.

When one party hangs up, the switches exchange the usual series of ISUP REL messages, followed by an RLC message. If the fixed-line PSTN/ISDN subscriber hung up first, the MSC sends a BSSAP DISC message to the MS when it receives the REL message; the MS should respond with a DTAP REL message. When the serving MSC receives the expected DTAP REL in return, it should finally release the connection by sending a DTAP REL_COM to the MS and an IAM REL through the GMSC back to the calling party's PSTN/ISDN switch. If the PLMN subscriber hung up first, the MS sends a DTAP DISC message to the serving MSC, which then initiates the ISUP REL and sends a DTAP REL back to the MS. The MS should respond with a DTAP REL_COM to confirm the release; this response allows the serving MSC to send an ISUP RLC back through the network to the calling party's PSTN/ISDN switch, thereby releasing the connection.

sendRoutingInfo (SRI)

In the case of a mobile terminating call, the GMSC sends this message to the called party's HLR to obtain routing information, such as the MSRN. Upon receiving the message, the HLR sends a provideRoamingNumber request to the VLR where the subscriber is currently roaming.

provideRoamingNumber (PRN)

The VLR uses this message to provide routing information (MSRN) to the HLR in the case of a mobile terminating call, which is sent to the GMSC. See Figure 13-7 and the description of sendRoutingInfo for more information.

In Appendix L, Example L-4 shows a trace that depicts an HLR decode calling a VLR to request an MSRN using the provideRoamingNumber operation. Also in Appendix L, Example L-5 shows how a trace illustrates a VLR's decode calling an HLR to return an MSRN that uses the provideRoamingNumber operation.

Supplementary Services

Supplementary services include the following operations:

- registerSS
- eraseSS
- activateSS
- deactivateSS
- interrogateSS
- registerPassword
- getPassword

In addition to these supplementary services, the following operations are considered unstructured supplementary services:

- processUnstructuredSS-Request
- unstructuredSS-Request
- unstructuredSS-Notify

The following section introduces the unstructured supplementary services (USSs) concept and discusses operations.

Unstructured Supplementary Services (USSs)

GSM 02.04 defines supplementary services. In addition to supplementary services, GSM has defined the concept of USSs. USSs allow PLMN operators to define operator-specific supplementary services and to deliver them to market quickly. The final three operations listed at the beginning of this chapter are used in USS implementation. USS allows the MS (subscriber) and the PLMN operator-defined application to communicate in a way that is transparent to the MS and intermediate network entities.

The communication is carried out using Unstructured supplementary service data (USSD) data packets, which have a length of 80 octets (91 ASCII characters coded, using seven bits) and are carried within the MAP operation. USSD uses the dialogue facility (which is connection oriented) of TCAP and is specified in GSM 02.90 (USSD Stage 1) and GSM 03.90 (USSD Stage 2). Unlike SMS, which is based on a store and forward mechanism, USSD is session oriented and, therefore, has a faster turnaround and response time than SMS, which is particularly beneficial for interactive applications. USSD can carry out the same two-way transaction up to seven times more quickly than SMS can.

The wireless application protocol (WAP) supports USSD as a bearer; the mobile chatting service relies on USSD transport for the text, and most, if not all, prepay roaming solutions are implemented using USSD. With such prepay applications, the subscriber indicates to the network from a menu on the MS the desire to place a roaming call. The serving MSC connects to the subscriber's HLR, which sends the request to a USSD gateway, which, in

turn, sends the request to a prepay application server. The server checks the balance and then issues call handling instructions back to the MSC in the visited network. USS is still likely to find applications even in 3G networks.

Operations

The following bullets describe the operations for supplementary services and unstructured supplementary services:

- registerSS

The registerSS operation is used to register a supplementary service for a particular subscriber. The supplementary service (such as call forwarding) is often automatically activated at the same time.

- eraseSS

EraseSS is used to delete a supplementary service that was entered for a particular subscriber using registerSS.

- activateSS

ActivateSS is used to activate a supplementary service for a particular subscriber. Example supplementary services include CLIP/CLIR.

- deactivateSS

This operation switches off a supplementary service for a particular subscriber; it is the reverse of activateSS.

- interrogateSS

InterogateSS allows the state of a single supplementary service to be queried for a particular subscriber in the HLR.

- registerPassword

This operation is used to create or change a password for a supplementary service. When the HLR receives this message, it responds with a getPassword message to request the old password, the new password, and a verification of the new password. If the old password is entered incorrectly three consecutive times, this operation is blocked.

- getPassword

The HLR sends this message if the subscriber wants to change his current password or modify or activate a supplementary service. See also registerPassword. This operation is blocked if the old password is entered incorrectly three consecutive times.

- processUnstructuredSS-Request

This message is used to provide a means to support non-GSM standardized supplementary services. Both the MS and the addressed NSS network entity use it, only if the MS initiated the transaction.

- unstructuredSS-Request

Same as processUnstructuredSS-Request, except that both the MS and the addressed NSS network entity use it, only if the NSS entity initiated the transaction.

Short Message Service (SMS)

SMS provides paging functionality for alphanumeric messages of up to 160 characters to be exchanged with other GSM users. The network itself can also generate messages and broadcast to multiple MSs or to a specific MS. For example, a welcome message can be sent to a subscriber when he or she roams onto a new network; in addition, it can provide useful information, such as how to retrieve voicemail. The SMS service also transfers ring tones and logos to the MS.

The SMS slightly blurs the image of the user traffic being separate from signaling because, in a sense, the messages are user traffic; they are for human processing (written and read), rather than for communication between network entities.

The SMS does not have subcategories. It has the following operations:

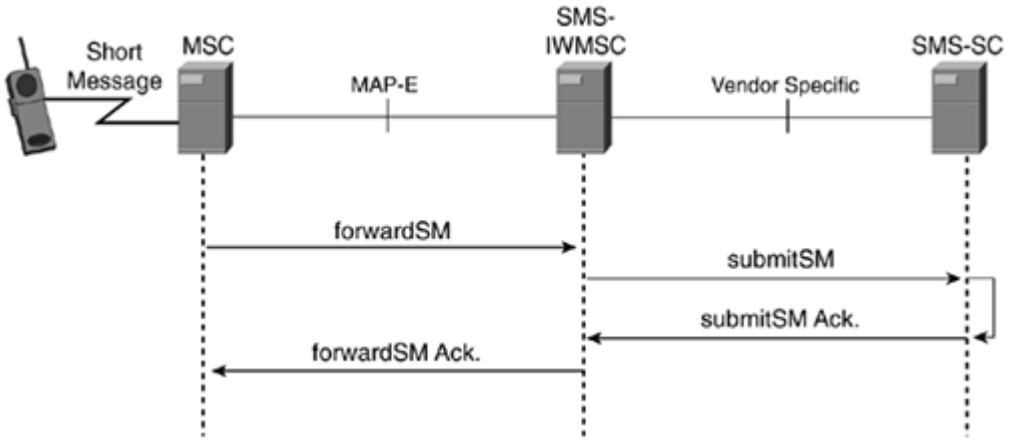
- forwardSM
- sendRoutingInfoForSM
- reportSMDeliveryStatus
- readyForSM
- alertServiceCentre
- informServiceCentre

The following sections examine each of these.

forwardSM

Both the mobile originating (MO-SMS) and mobile terminating SMS (MT-SMS) procedures use the forwardSM operation to carry text messages between the MSC where the subscriber roams and the SMS-IWMSC or the SMS-GMSC, respectively. Figure 13-8 shows the MO-SMS procedure.

Figure 13-8. MAP Operations Involved in Sending an SMS from MS to the SMS-SC



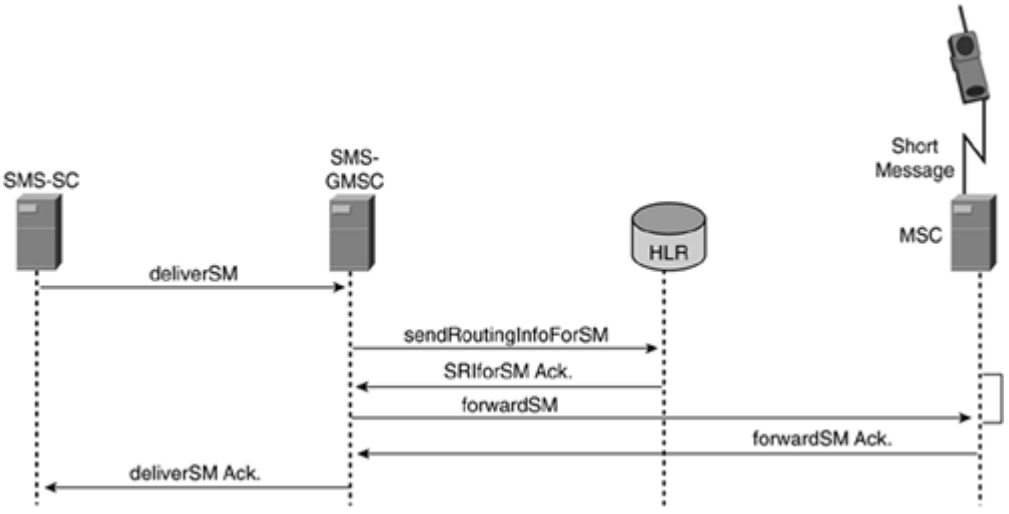
In Appendix L, Example L-6 contains a trace that shows the decode of a MAP operation forwardSM, including its SMS text.

sendRoutingInfoForSM

The SMS-GMSC uses this message during an MT-SMS to deliver an SMS to the MSC in whose area the subscriber is currently roaming. The message contains the subscriber's MSISDN, and the result contains the destination MSC's ISDN number. SCCP then uses this ISDN number to deliver the SMS using a forwardSM message. Figure 13-9 shows the MT-SMS procedure.

Figure 13-9. MAP Operations Involved in Sending an SMS from the SMS-SC to the MS

[View full size image]



In Appendix L, Example L-2 shows a trace showing a VLR's decode calling an HLR (to perform a location update).

reportSMDeliveryStatus

If the SMS-SC cannot deliver the MT-SMS to the MS (because the subscriber is not reachable, for example), then the SMS-SC returns a negative result to the SMS-GMSC. Upon receiving this result, the SMS-GMSC sends a reportSMDeliveryStatus to the HLR, which, in turn, sets a message waiting flag in the appropriate subscriber data. The HLR also sends an alertServiceCentre message to the SMS-IWMSC to inform it about the negative SM delivery and waits until the subscriber can be reached. When the VLR (also aware of SM delivery failure) detects that the subscriber is again reachable, it sends a readforSM message to the HLR. The HLR, in turn, sends an alertServiceCentre message to the SMS-IWMSC, which informs the SMS-SC. The delivery process then begins again with a forwardSM message.

NOTE

The previous section also pertains to the readyForSM and alertServiceCentre.

informServiceCentre

If a sendRoutingInfoForSM is received for a subscriber that is currently unavailable, the HLR sends this message to the SMS-GMSC.

Summary

MAP primary use is to allow calls to be delivered to mobile subscribers. Unlike with fixed-line networks, the subscriber's location cannot be determined from the numbering scheme that is used in the network. Therefore, the subscriber's location must be known in real-time so a call can be connected to the nearest switch to the mobile subscriber. MAP keeps track of a mobile subscriber and provides other functionality, including allowing mobile subscribers to send alphanumeric two-way text between handsets; this is known as SMS. MAP also provides mobile operator's with the functionality to manage a subscriber's subscription so that services can be added and removed in real-time.