ACKNOWLEDGMENT

First of all I would like to thank God for guiding me through my study.

More over I want to pay special regards to my parents who are enduring these all expenses and supporting me in all events. I am nothing without their prayers. They also encouraged me in crises. I shall never forget their sacrifices for my education so that I can enjoy my successful life as they are expecting.

Also, I feel proud to pay my special regards to my project adviser "Dr. Jamal Fathi". He never disappointed me in any affair. He delivered me too much information and did his best of efforts to make me able to complete my project. I would like to thank Assoc. Prof. Dr. Adnan Khashman for his advices in each stag of our undergraduate program and how to choose the right path in life.

The best of acknowledge, I want to honor those all persons who have supported me or helped me in my project. I also pay my special thanks to my all friends who have helped me and gave me their precious time to complete this project. Also my especial thanks go to my friends, Ahmad Ibrahim, Mahmoud Al-Qassas and

Eng. Ibrahim Abu-awwad.

ABSTRACT

Signaling System 7 (SS7) is a global standard that defines the architecture and protocol used by Public Switched Telephone Networks (PSTN). Call Setup, call forwarding, voice mail, toll free calling, and customer billing are some of the functions of SS7. There are many different carriers providing these services. Each carrier wants to provide a high quality of service for the customer and generate revenue. To provide quality of service the carrier must ensure calls are error free. The carrier also needs to provide security to detect fraud or attacks. The carrier can protect the network by gathering statistics and monitoring for abnormal patterns. This project gives an overview of the SS7 architecture and protocol and describes the importance of monitoring the SS7 network.

CONTENTS

INTRODUCTION

In the last hundred years telephones have become a standard item in any home, office and on the street. No one gives a second thought to the action of picking up a receiver, dialing a number, and hearing a voice at the other end, with all the actions occurring within less than a minute. Very few people realize the flurry of activities occurring in that minute. In the early days of telephony you would first call an operator sitting at the switchboard, give the operator the telephone number of the called party and upon determining that the other end is free and available the operator would complete the connection, allowing you to talk.

The process of establishing connections became automated, and in the place of a switchboard, we now have electronic switches. Switches communicate with each other using signaling messages traveling through the network. There are two types of signaling: in-band and out-of-band.

Signaling System 7 (SS7) is the currently used standard for the telephone signaling network. It uses common channel signaling, which is a "signaling method in which a single channel conveys, by means of labeled messages, signaling information relating to a multiplicity of circuits or calls and other information, such as that used for network management [3]." It means that a separate network with its own nodes and links was built to provide support to the conventional voice network. SS7 is a digital packet switched network that can carry information about a number of calls over the same link simultaneously. It is responsible for connection set up, control and tear down, as well as routing and network maintenance. With SS7, signaling can take place during the conversation instead of only at the beginning.

The goals of this project are to understand the need of signaling system No. 7, and how the messages in SS7 network are transferred in order to maintain a connection between users with all the services that SS7 can provide.

The first chapter of this project provides an introduction to telephone networks and their main subsystems. Types of transmitting, signaling and switching techniques are discussed and a quick explanation of how a telephone call is made.

Chapter two covers the different methods of signaling systems, with some information about the functions of signaling. I defined common channel signaling system which is the backbone to the revolution of signaling system No. 7.

Chapter three provides a detailed overview of the SS7 architecture. I viewed the main nodes and links of SS7 network. I viewed the protocol stacks of SS7 networks and a comprehensive explanation of each layer was given. And I talked about the signaling units that used in the transmission of the SS7 messages over the network.

Chapter four describes various SS7 applications in the telecommunication networks, providing many services and benefits to bridge the users all over the world.

And finally, chapter five contains an overview of VoIP networks and the interworking between this technology and SS7. This technology still evolving and can introduce many new applications in the near future.

1. INTRODUCTION TO TELEPHONE NETWORKS

1.1 Overview

In order to get a telephone call to travel from one place to another, it must pass through the telephone network. This network consists of many different parts, operated by many different companies, but is inter-connected using common signaling methods. This chapter will discuss the basics of the telephone network, different methods of transmission, switching, signaling between different telephone systems, and the use of tandems (transfer) in the network.

1.2 The Telephone Network

The public-switched telephone network (PSTN) consists of transmission components, switching components, and facilities for maintenance equipment, billing systems, and other internal components. Transmission components (links) define the cable or wireless infrastructure for transmitting signals. Switching components (nodes) include transmitters and receivers for setting up voice circuits.

1.2.1 Nodes and Links

A network is comprised of two fundamental parts, the nodes and the links. A node is some type of network device, such as a computer. Nodes are able to communicate with other nodes through links, like cables. There are basically two different network techniques for establishing communication between nodes on a network: the circuit-switched network and the packet-switched network techniques [4]. The former is used in a traditional telephone system, while the latter is used in IP-based networks. Figure 1.1 shows some types of nodes and links.

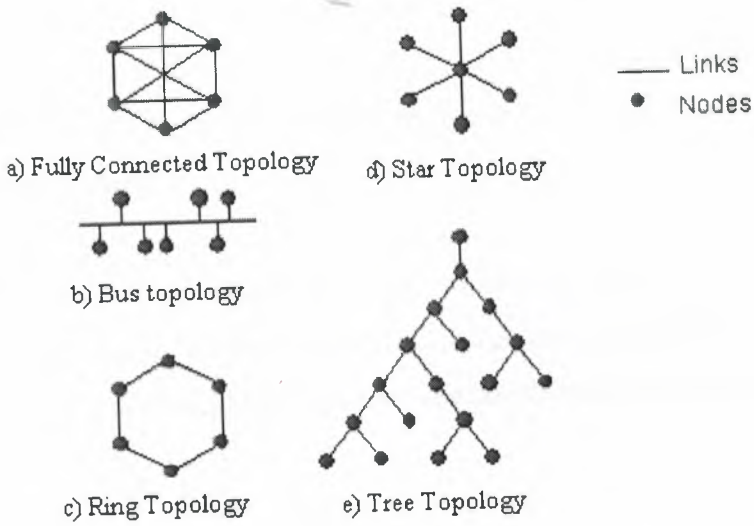

Figure 1.1 Types of Nodes and Links

1.2.2 Hierarchy

The telephone system was originally designed as a hierarchy of switches that set up calls across COs, across LATAs, or across long-distance connections. This hierarchy is pictured in Figure 1.2.

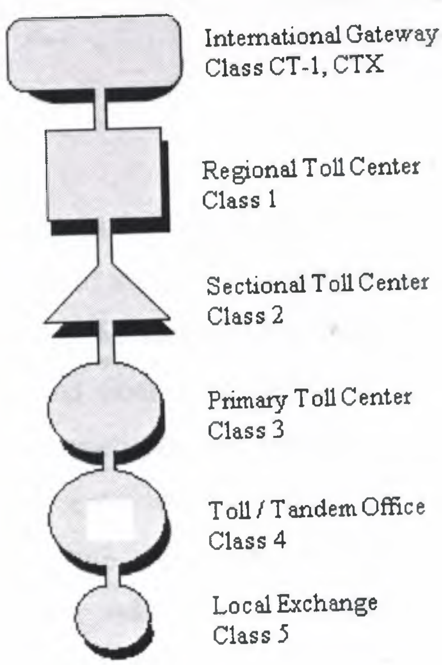

Figure 1.2 Pre-Divestiture Switching Hierarchies

Long distance calls entering the network through the caller's end office would climb the switching hierarchy in search of an idle circuit. If the most direct route was busy, the call moved up the hierarchy to the next switching center, and the next, until a path was found to complete the call.

The Class 1 office is the highest level to which the search can be carried out. International Gateways are equipped with software that can translate telecommunications protocols used in one portion of the world into protocols recognizable by switches at the destination.

1.3 The Network Subsystems

The most fundamental principle of that network is quite simple; because it would be impossible to install a line from each caller to every other caller, the telephone system is a switched network. A switched network brings each subscriber line into a centralized switching system, where connections are made for each call.

In addition to the switching systems that route incoming voice and data calls to their destinations, the two other key components of a public switched telephony network are transmission and signaling.

1.3.1 Transmission

The transmission segment of telephony networks is concerned with moving information from one location to another. Transmission can use either analog or digital signals, and those signals can be carried over various transmission media, such as copper wire, radio waves, and fiber-optic cable. For companies seeking to establish differentiators from their competitors, the transmission technologies have taken on increasing importance in recent years as digital and fiber-optic technologies have provided network providers with the means to deliver exciting new high bandwidth services, such as ISDN, dialable video, or other wideband services [3].

1.3.1.1 Analog Transmission

While some analog transmission systems are still active in the public network, most providers are moving as quickly as possible to digital technology. All transmission carried over long distances must be amplified periodically because the signal is always losing strength and unwanted "noise" is always being introduced into the signal. Analog transmission systems have severe limitations in this amplification process. Analog amplifiers not only amplify the voice and data signals, but also the noise. And because analog waveforms can vary continuously over the bandwidth, much of the original signal can be lost because it is very difficult for the receiving end to exactly reproduce the original waveform from the distorted, noisy transmission.

Analog transmission is sufficient for most voice transmissions, because a small inaccuracy in the received signal will not be detected by a listener. But accurate transmission is absolutely essential to data transmission, where a single changed bit could completely destroy the meaning of the original signal [3].

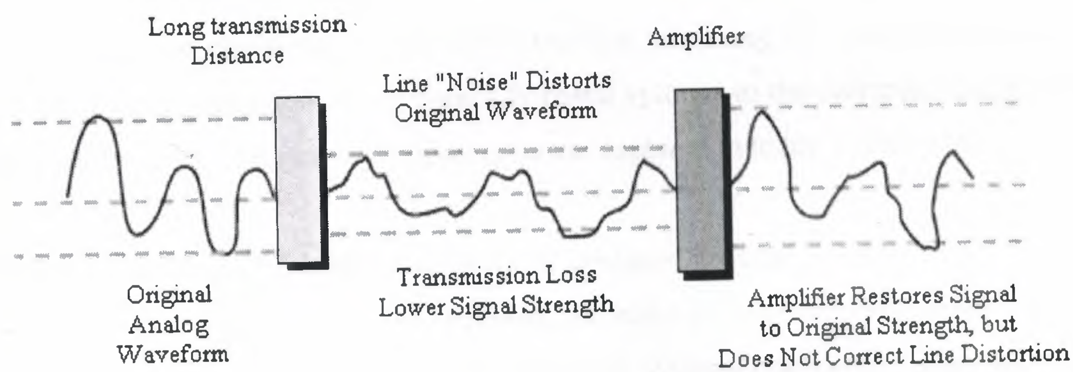

Figure 1.3 Amplifier for Analog Transmission System

1.3.1.2 Digital Transmission

Digital signals can be transmitted over great distances and coded, regenerated, and decoded with no degradation of quality. Digital transmission employs repeaters rather than amplifiers. Repeaters read the incoming signal (which has been distorted and weakened during transmission) and determines the original sequence of discrete signal levels. The repeater then reproduces the original signal and sends it to the next network destination. Coupled with fiber-optic transmission gear which transmits the 0s and 1s through bursts of light only a digital network can handle high-speed data and graphics/video transmission, as well as voice calls.

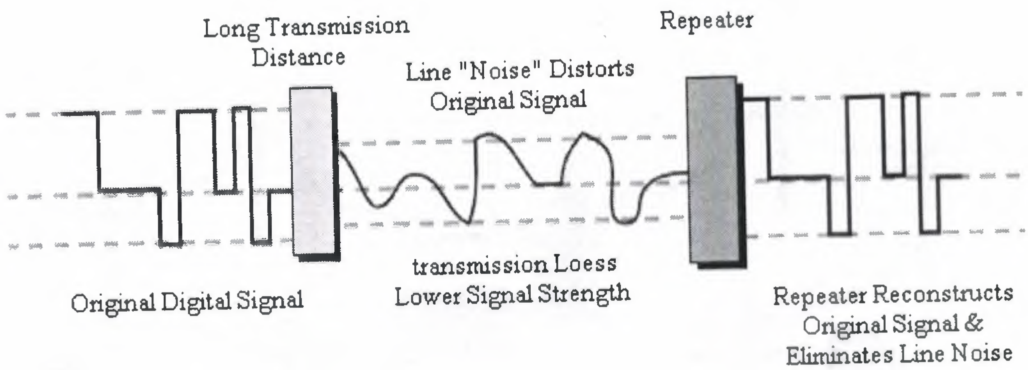

Figure 1.4 Repeater for Digital Transmission System

1.3.2 Switching

The history of telephony switching has exactly mapped the introduction and development of electronics technologies into the telephony life over the last century, from the early perfection of mechanically-based systems to the dramatic changes made possible by the abundance of computing power captured in today's silicon [4].

The public switched telephone network is designed so that any user can be directly connected to any other user on the network. To make these connections economical and practical, switching systems are used. Switching systems concentrate many users onto relatively few distribution paths and provide a connection over the distribution path to the called party. The terms concentration, distribution and expansion relate to functions that must be performed in order to connect any inlet over a path or any outlet.

1.3.2.1 Manual Switching

The earliest telephone switches were manual, that is, they required a human operator to make connections by plugging circuits into a switchboard. When a customer "rang" the central office, the operator scanned the switchboard and connected the caller by plugging into the requested line.

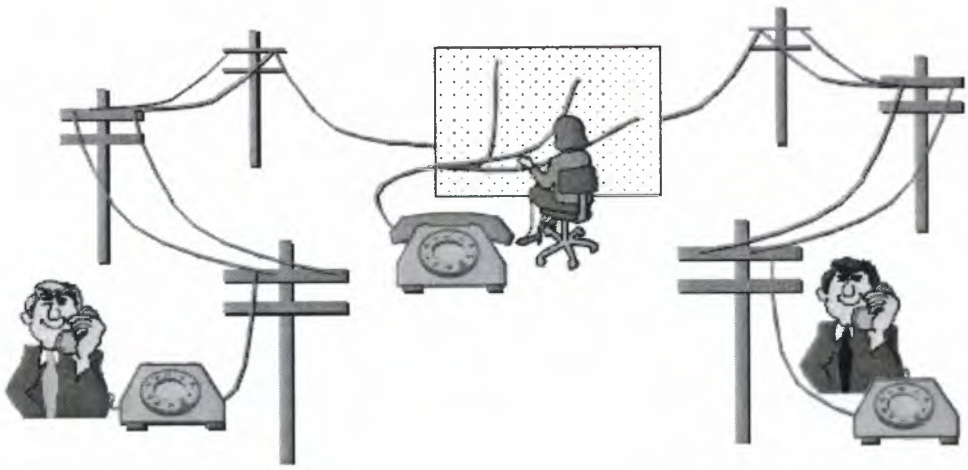

Figure 1.5 the Process of Manual Switching

1.3.2.2 The Strowger Switch

The Strowger "step-by-step" switch, like all early systems, was based on the analog technology that was state-of-the-art electronics at the time. The major drawbacks of Strowger switching were the large amount of space it occupied and the high electrical power consumption needed during busy-hour operation. And because the mechanical parts were subject to wear and electrical contacts were sensitive to damage and dirt, maintenance for a Strowger switch was extremely hard and labor-intensive.

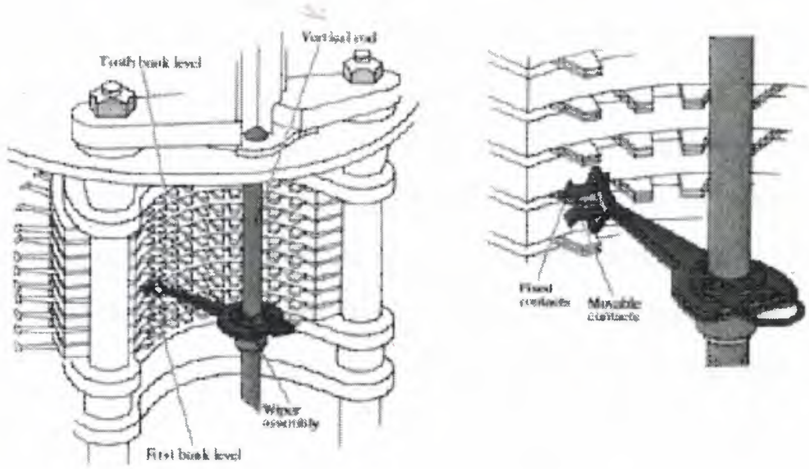

Figure 1.6 the Strowger Switch

1.3.2.3 Digital Switching

In the mid 1970s Telecom companies began introducing digital technologies into the core of the public switched network. Digital switches fully capitalize on the strength of the computer revolution by routing both voice and data through the switch in the form of 0/1 binary coded information, which can be moved through the switch in a very short period of time. Because the digital switch was faster, smaller, more able to efficiently handle data, and provided infinitely more bang for the buck especially when power, real estate, and maintenance costs were factored in the digital switch soon became the standard switching system in North America. A single digital switch typically serves anywhere from under 1000 to over 100,000 subscribers.

1. How Digital Technology Works

The familiar telephone set creates an analog wave representation of the human voice by using the air pressure from speech to vibrate the diaphragm in the handset, which in turn activates carbon granules to produce an electrical signal of varying strengths.

2. Pulse Code Modulation - Converting Analog to Digital Signals

In a process called pulse code modulation (PCM), this analog wave signal is then sampled every 125 microseconds (8000 times a second) and converted to pulses. The amplitude of each pulse is represented by the amplitude of the analog signal. This PCM signal is then put into 8-bit packets and "quantized" by measuring and rounding it off to one of 255 levels of sound. Each sample is then converted into a digital representation of the PCM packet that is, it is converted into a code represented totally by discrete 0s and 1s.

3. Frequency-Division Multiplexing (FDM)

Frequency-division multiplexing (FDM) is a form of signal multiplexing where multiple baseband signals are modulated on different frequency carrier waves and added together to create a composite signal [1].

Historically, telephone networks used FDM to carry several voice channels on a single physical circuit. In this, 12 voice channels would be modulated onto carriers spaced 4 kHz apart. The composite signal, occupying the frequency range 60 – 108 kHz, was known as a group. In turn, five groups could themselves be multiplexed by a similar method into a supergroup, containing 60 voice channels. There were even higher levels of multiplexing, and it became possible to send thousands of voice channels down a single circuit. Modern telephone systems employ digital transmission, in which time-division multiplexing TDM is used instead of FDM.

4. Time Division Multiplexing (TDM)

In a key process known as time division multiplexing, these 8-bit digital signals (each a "channel" of information) are then combined into a 24 channel, 125 microsecond frame. This multiplexing allows many channels of information to be simultaneously transmitted over the same pathway, as pieces of the signal are woven together one after the other and assigned time slots on the pathway. In most North American digital systems, 24 channels (i.e., 24 voice or data "conversations") are simultaneously carried on a single pathway. An often cited analogy of TDM is to imagine a train carrying 24 conversations, with a piece of each conversation in each of its boxcars. As the train rolls up to the loading dock, in figure 1.7, conversation 1 is put into boxcars 1, 25, 49, etc. Conversation 2 is put into boxcars 2, 26, 50, etc. and so forth. At the destination, the process is reversed and the conversation is reassembled. In a similar fashion digital signals travel through the network in groups of 24 multiplexed channels.

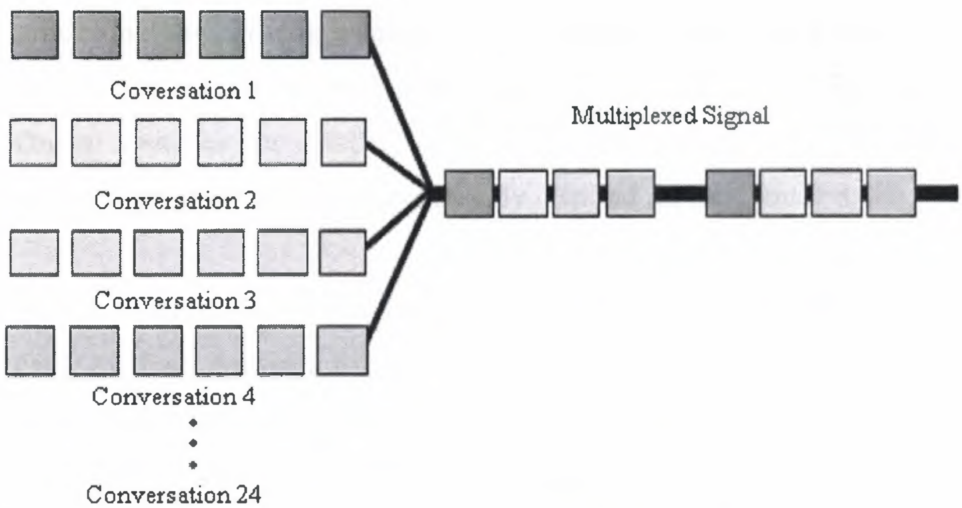

Figure 1.7 Time Division Multiplexing

5. **Time Division Switching in the Digital Switch**

Digital switches use TDM to process a huge number of calls in the smallest amount of time. In the DMS SuperNode and DMS-10 400 Series systems, for instance, digital signals are multiplexed onto 32-channel links (called time slots) to the switching matrix of the switch. The switching matrix uses time division switching to place this incoming traffic onto the proper outgoing time slots to lines and trunks. For example, in the figure 1.7, time slot 1 is mapped to time slot 4 to make the proper connection. After switching, the digital signals are multiplexed back together and sent to the called party.

The typical digital switch has four essential components: the central processor, the switch matrix, a range of peripherals, and input/output controllers. And I will give a brief explanation for each one of them.

- **Central Processor:** The central processor controls call processing activities for example, assigning time slots and administering features such as call forwarding, as well as directing system-control functions, system maintenance, and the loading and downloading of software. To ensure reliability, the central processor is generally duplicated on digital switches. Each call is processed simultaneously on both processors; if the "hot" processor should develop a problem, the system automatically shifts to the standby processor without the caller noticing any interruption of service. State-of-the-art larger digital switches are increasing processing power through additional specialized processors for functions such as frame-relay data, ISDN packet handling, service control point functionality, etc. Digital switches for smaller, rural communities often adopt alternate service access strategies that can be more easily justified for these markets.

- **Switch Matrix**: also referred to as the network, handles the actual connection of calls to their destinations. The latest switching modules, such as the DMS SuperNode Enhanced Network (ENET), can process up to 64,000 channels in a single cabinet and switch wideband data as effortlessly as a voice conversation.

- **Peripherals**: The typical digital switch has a range of peripheral modules to interface the range of lines and trunks coming in from the network. The peripherals convert incoming voice and data signals into the digital format used by the switch and perform some low-level call processing tasks. Typical peripherals include those that terminate lines, trunks, digital loop carriers, and maintenance trunks.

- **Input/Output Controller system**: provides access to the switch for maintenance, billing, routine operations and administration, and loading of software.

6. Call Processing

The instant appearance of a dial tone is an unquestioned expectation of most telephone users in world wide. What the caller does not see is the stream of instructions executed by the switch before dial tone occurs. Dial tone, in fact, is delivered, not from the telephone itself, but from a central office switch located perhaps miles away. By the time the dialing sequence is complete, the switch has performed thousands of "transparent" call processing activities.

The primary function of the switch is to establish connections between telephones and data equipment for the transmission of voice or data. When a local call is placed, the fundamental switch call processing components come into play. These are [4]:

- **Call Detection -** Detecting that the telephone receiver has been lifted.
- **Dial Tone Provided -** Providing a dial tone to the caller.
- **Digit Collection -** Collecting the dialed digits.
- **Digit Translation -** Translating the digits dialed to a called number.
- **Call Routing -** Routing the call to its destination.
- **Call Connection -** Establishing a connection between the parties.
- **Audible Ringing/Ringback -** Signaling the called party by audible ringing, and the calling party by ringback.
- **Speech Path Established -** Detecting when the called line is answered.
- **Call Termination -** Detecting disconnect and terminating the call when a party hangs up.

When the handset rests on the cradle, the circuit is on-hook. In other words, before a phone call is initiated, the telephone set is in a ready condition waiting for a caller to pick up its handset. This state is called on-hook. In this state, the -48VDC circuit from the telephone set to the CO switch is open. The CO switch contains the power supply for this DC circuit. The power supply located at the CO switch prevents a loss of telephone service when the power goes out at the location of the telephone set. Only the ringer is active when the telephone is in this position [5]. Figure 1.8 shows the off-hook phase.

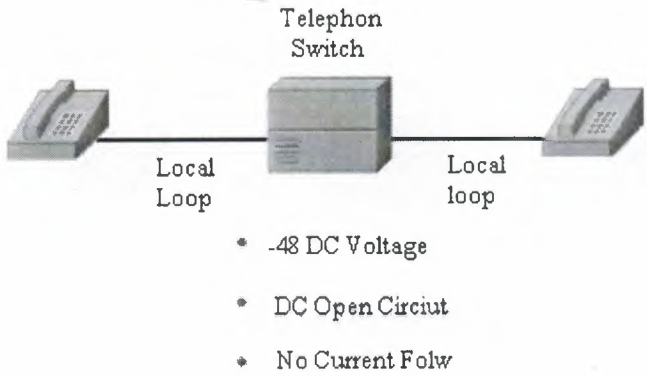

Figure 1.8 On-hook Call Progress

The off-hook phase occurs when the telephone customer decides to make a phone call and lifts the handset from the telephone cradle. The switch hook closes the loop between the CO switch and the telephone set and allows current to flow. The CO switch detects this current flow and transmits a dial tone (350- and 440-hertz [Hz] tones played continuously) to the telephone set. This dial tone signals the customer can begin to dial. There is no guarantee that the customer hears a dial tone right away. If all the circuits are used, the customer could have to wait for a dial tone.

The access capacity of the CO switch used determines how soon a dial tone is sent to the caller phone. The CO switch generates a dial tone only after the switch has reserved registers to store the incoming address. Therefore, the customer cannot dial until a dial tone is received. If there is no dial tone, then the registers are not available. Figure 1.9 shows the dialing phase.

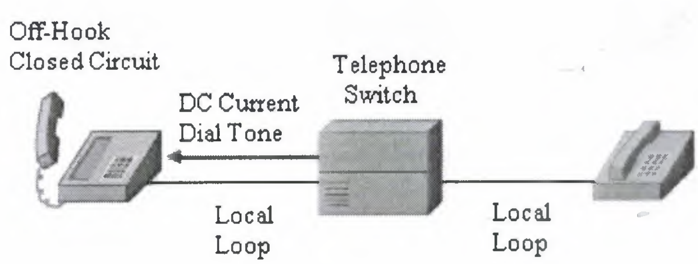

Figure1.9 Off-Hook Call Progress

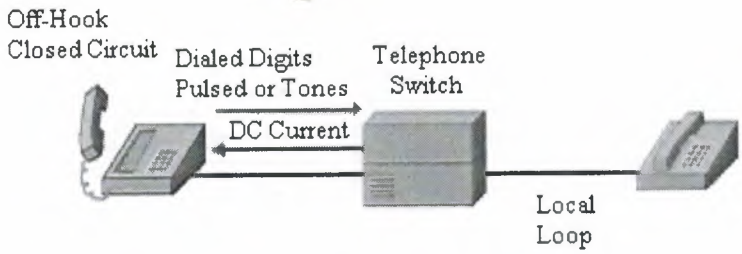

Figure 1.10 Dialing Phase

The dialing phase allows the customer to enter a phone number (address) of a telephone at another location. The customer enters this number with either a rotary phone that generates pulses or a touch-tone (push-button) phone that generates tones. These telephones use two different types of address signaling in order to notify the telephone company where a subscriber calls: Dual tone multi frequency (DTMF) dialing and Pulse dialing.

These pulses or tones are transmitted to the CO switch across a two-wire twisted-pair cable (tip and ring lines). Figure 1.11 shows the switching phase.

In the switching phase, the CO switch translates the pulses or tones into a port address that connects to the telephone set of the called party. This connection could go directly to the requested telephone set (for local calls) or go through another switch or several switches (for long-distance calls) before it reach its final destination. Figure1.12 shows the ringing phase.

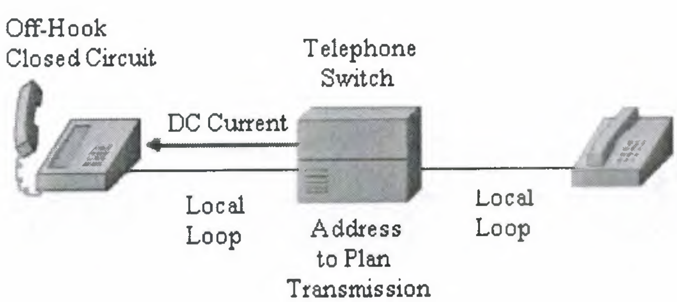

Figure 1.11 Switching phase

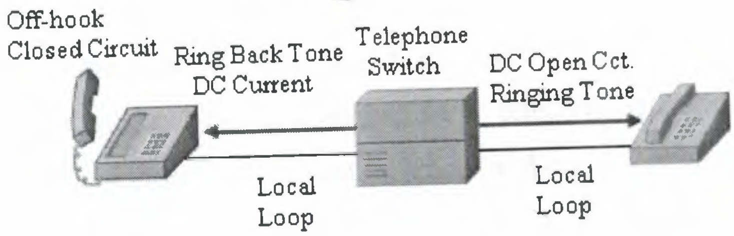

Figure 1.12 Ringing Phase

Once the CO switch connects to the called line, the switch sends a 20-Hz 90V signal to this line. This signal rings the phone of the called party. While ringing the phone of the called party, the CO switch sends an audible ring-back tone to the caller. This ring-back lets the caller know that ringing occurs at the called party. The CO switch transmits 440 and 480 tones to the caller phone in order to generate a ring-back. These tones are played for a specific on time and off time. If the called party phone is busy, the CO switch sends a busy signal to the caller. This busy signal consists of 480- and 620-Hz tones.

In the talking phase, the called party hears the phone ringing and decides to answer. As soon as the called party lifts the handset, an off-hook phase starts again, this time on the opposite end of the network. The local loop is closed on the called party side, so current starts to flow to the CO switch. This switch detects current flow and completes the voice connection back to the calling party phone. Now, voice communication can start between both ends of this connection. Figure 1.13 shows the talking phase.

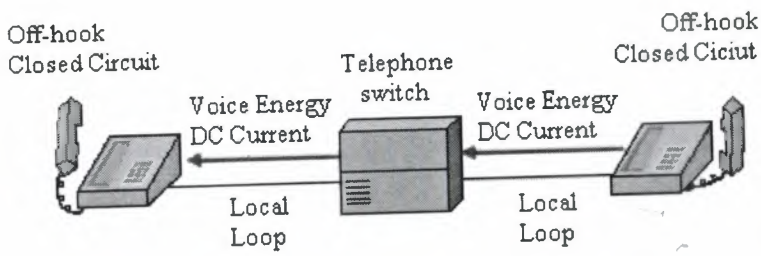

Figure 1.13 Talking Phase

1.3.3 Signaling

In telecommunication, the term "signaling" has the following meaning of the use of signals for controlling communications. In a telecommunications network, the information exchange concerning the establishment and control of a connection and the management of the network, in contrast to user information transfer. The sending of a signal from the transmitting end of a circuit to inform a user at the receiving end that a message is to be sent [2].

1.4 Summary

In this introductory chapter, we reviewed the technology of telephone networks and their main building blocks. Then we presented the key concepts in telephony transmission, switching and how digital switching controls the process of any call in the present telephone networks.

Next chapter will present detailed information about the structure and principles of operation of signaling systems used in telecommunication networks.

2. SIGNALING

2.1 Overview

The telephone network is a widely distributed system of intelligent switching nodes. Signaling is the process by which nodes communicate to establish and tear down calls so that two or more parties can communicate via terminal equipment (such as a phone acting as an endpoint) and the network. This chapter discusses the signaling techniques required to control voice transmission.

2.2 Need for Signaling

A network, whether public or composed of a group of privately interconnected nodes, would be of limited use unless users are able to communicate their needs for service. In addition to this user–network signaling, signaling capabilities between various components of the network are needed. Signaling system and related equipment are used by all public or private telecommunication networks, with the exception of some types of data communication networks, which use its own mechanisms [2].

For traditional telecommunication services over the public switched network or over a private voice network, signaling refers to the mechanism necessary to establish a connection, to monitor and to supervise its status, and to terminate it, through the transmission and switching fabric of the underlying network. Formally, signals are message generated by the user of some internal network processor, pertaining to call management. Signaling is the act of transferring this information among remote entities, including the communication handshake protocols and the "semantic" conventions. The signaling network is the collection of physical transport facilities that carry the signals. The signaling equipment performs the functions of alerting, addressing, supervising, and providing status in communication carrier networks.

2.3 Functions of Signaling

Signaling events occurring during network call processing may be divided into two basic categories: address signaling and supervisory signaling.

2.3.1 Address Signaling

Address signaling is the means by which a subscriber or switching system inputs dialed number information into the network. In some applications, the telephone equipment outputs automatic number identification (ANI) information into the network following an off-hook. Typically address signaling is accomplished by dial pulsing or by in-band signaling with DTMF and MF tones. This information must often be transmitted over several links in the switched network to establish a voice path between the caller and the called party.

2.3.1.1 Pulse Dialing

Pulse Dialing is an in-band signaling technique. It is used in analog telephones that have a rotary dialing switch. The large numeric dial-wheel on a rotary-dial telephone spins to send digits to place a call. These digits must be produced at a specific rate and within a certain level of tolerance. Each pulse consists of a "break" and a "make", which are achieved when the local loop circuit is opened and closed. The break segment is the time during which the circuit is open. The make segment is the time during which the circuit is closed. Each time the dial is turned; the bottom of the dial closes and opens the circuit leading to the CO switch or PBX switch.

A "governor" inside the dial controls the rate at which the digits are pulsed; for example, when a subscriber dials a digit on the rotary dial to call someone, a spring winds. When the dial is released, the spring rotates the dial back to its original position, and a cam-driven switch opens and closes the connection to the telephone company. The number of consecutive opens and closes--or breaks and makes--represents the dialed digits Therefore, if the digit 3 is dialed; the switch is closed and opened three times. Figure 2.1 represents the sequence of pulses that occur when a digit 3 is dialed with pulse dialing.

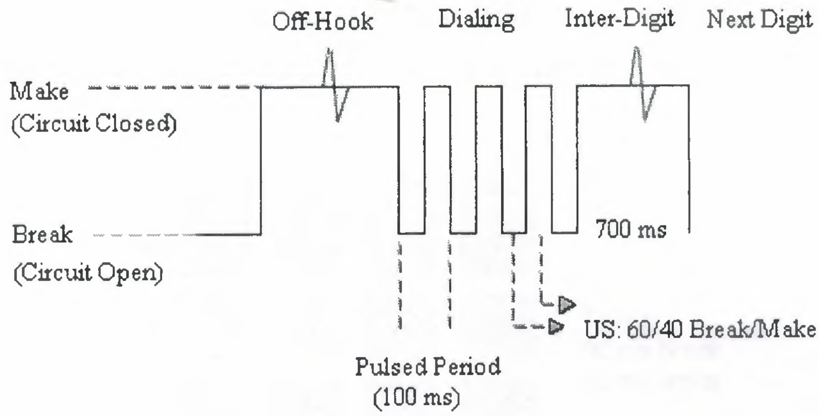

Figure 2.1 Pulse Dialing

This illustration displays the two terms, make and break. When the telephone is off-hook, a make occurs and the caller receives a dial tone from the CO switch. Then the caller dials digits, which generate sequences of makes and breaks that occur every 100 milliseconds (ms). The break and make cycle must correspond to a ratio of 60 percent break to 40 percent make. Then the phone stays in a make state until another digit is dialed or the phone is put back to an on-hook (equivalent to a break) state. Dial pulse addressing is a very slow process because the number of pulses generated equates to the digit dialed. So, when a digit 9 is dialed, it generates nine make and break pulses. A digit 0 generates ten make and break pulses. In order to increase the speed of dialing, a new dialing technique (DTMF) was developed.

2.3.1.2 Dual Tone Multi-Frequency (DTMF)

DTMF dialing is an in-band signaling technique just like pulse dialing. This technique is used in analog telephone sets that have a touch-tone pad. This dialing technique uses only two frequency tones per digit, as shown in Figure 2.2. Each button on the keypad of a touch-tone pad or a push-button telephone is associated with a set of high and low frequencies. On the keypad, each row of the key is identified by a low-frequency tone, and each column is associated with a high-frequency tone. The combination of both tones notifies the telephone company of the number called, hence the term dual tone multi-frequency. Therefore, when digit 0 is dialed, only frequency tones 941 and 1336 are generated instead of the ten make and break pulses generated by pulse dialing. The timing is still a 60-ms break and 40-ms make for each frequency generated. These frequencies were selected for DTMF dialing based on their insusceptibility to normal background noise.

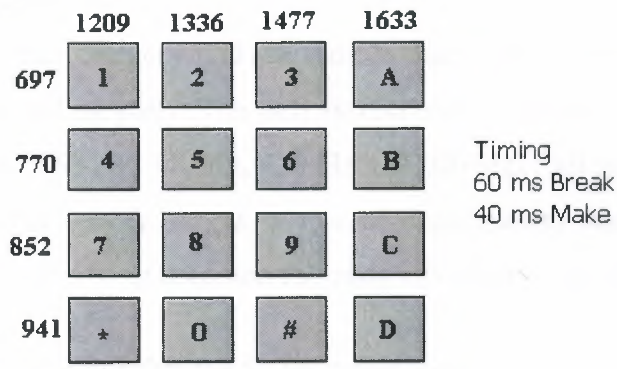

Figure 2.2: Dual Tone Multi-Frequency

2.3.2 Call Supervision

Call supervision detects or changes the state or condition of a line or trunk (via out-of-band signaling). There are two possible supervised conditions: on-hook and off-hook. On-hook means telephone equipment is idle; off-hook occurs when telephone equipment is active [5].

When a line/trunk goes off-hook, it is interpreted as a seizure by the system, and the line/trunk's operating state goes from idle to active. Both ends of a voice path must be off-hook for two-way communication to occur. If one end of the path goes on-hook and the other remains off-hook, the voice path becomes unidirectional (near- or far-end disconnect). The calling party can input control information to dial another number. When both ends of the path go on-hook, the voice path is torn down in the network.

2.3.2.1 Alert Signaling

The most familiar alert signal is power ringing, which notifies a called party of an incoming call. Ringing is initiated by applying ring voltage (50 to 130 V @ 20/30 Hz) to the line or trunk (out-of-band signaling). Ring voltage is normally obtained from a ring generator which is wired to the switching system. The far end central office provides audible ringback (an in-band call progress tone) toward the calling party to indicate that ring voltage has been applied to the circuit.

2.3.2.2 Call Progress Tones

Signals in this category include audible tones that indicate the progress of a telephone call to a calling party. The Bell system uses a Precise Tone Plan consisting of four frequencies: 350 Hz, 440 Hz, 480 Hz and 620 Hz. Call progress tones consist of these frequencies (either single or paired frequencies) and specific temporal patterns (cadences). The most common call progress tones are as follows:

- **Dial tone:** Indicates that the CO is ready to except digits from the subscriber. The dial tone can be removed from the line when the first digit is detected.

- **Busy tone**: Indicates that the called line has been reached but it is engaged in another call.

- **Reorder tone**: Indicates that the local switching paths to the calling office or equipment serving the customer are busy or that a toll circuit is not available.

- **Special Information Tones (SITs)**: Indicate special network conditions encountered in both the Local Exchange Carrier (LEC) and Inter-Exchange Carrier (IXC) networks.

- **Audible Ring back:** Returned to the calling party to indicate that the called line has been reached and ringing has started.

2.4 Signaling Techniques

2.4.1 In – Band Signaling

In the earliest days of the telephone network, signaling was provided by means of direct current (DC) between the telephone instrument and the operator. Direct current (DC) signaling is accomplished through the use of two electrical states called "on-hook" and "off-hook". A third state may sometimes exist as a transition state. In a dc signaling arrangement, the supervisory or address signaling is done by superimposing one or more dc states on the same conductors that are used for voice transmission. Currently, two types of per-circuit in-band trunk supervisory signaling methods are available: loop reverse battery (based on subscriber line signaling methods) and E&M signaling.

2.4.1.1 Loop Reverse Battery

This type is applicable to trunks that require call origination and seizure at only one end (one-way trunks, for example, direct outward dialing trunks). The method employs open and closure signals from the originating end, and reversal of battery ground from the terminating end. At the originating end, on-hook is indicated by an open circuit; off-hook by a bridge circuit. At the terminating end, on-hook is indicated by a ground on the tip lead and -48 V on the ring lead of the circuit. And off-hook is indicated by -48 V on the tip lead and ground on the ring lead [3].

2.4.1.2 E&M Signaling

Another signaling technique used mainly between PBXs or other network-to-network telephony switches is known as E&M. E&M signaling supports tie-line type facilities or signals between voice switches. Instead of superimposing both voice and signaling on the same wire, E&M uses separate paths, or leads, for each. E&M is commonly referred to as ear and mouth or receive and transmit. There are five types of E&M signaling, as well as two different wiring methods (two-wire and four-wire). Table 2.1 shows that several of the E&M signaling types are similar.

Table 2.1 E&M Signaling Types

Type	M-lead off-hook	M-lead on-hook	E-Lead Off-hook	E-Lead on-hook
I	Battery	Ground	Ground	Open
II	Battery	Open	Ground	Open
III	Loop Current	Ground	Ground	Open
IV	Ground	Open	Ground	Open
V	Ground	Open	Ground	Open
SSDC5	Earth On	Earth Off	Earth On	Earth Off

Four-wire E&M Type I signaling is actually a six-wire E&M signaling interface common in North America. One wire is the E-lead; the second wire is the M-lead, and the remaining two pairs of wires serve as the audio path. In this arrangement, the PBX supplies power, or battery, for both M- and E-leads.

Type II, III, and IV are eight-wire interfaces. One wire is the E-lead, the other wire is the M-lead. Two other wires are signal ground (SG) and signal battery (SB). In Type II, SG and SB are the return paths for the E-lead and M-lead, respectively.

Type V is another six-wire E&M signaling type and the most common E&M signaling form used outside of North America. In Type V, one wire is the E-lead and the other wire is the M-lead.

Similar to type V, SSDC5A differs in that on- and off-hook states are backward to allow for fail-safe operation. If the line breaks, the interface defaults to off-hook (busy). Of all the types, only types II and V are symmetrical (can be back-to-back with a crossover cable). SSDC5 is most often found in England. The Cisco 2600/3600 series currently supports types I, II, III, and V utilizing both two- and four-wire implementations. This illustration depicts two-wire and four-wire E&M signaling connections. Voice travels over the tip and ring lines. Signaling occurs over E&M lines. Figure 2.3 illustrates type 1 E&M signaling with a two-wire line.

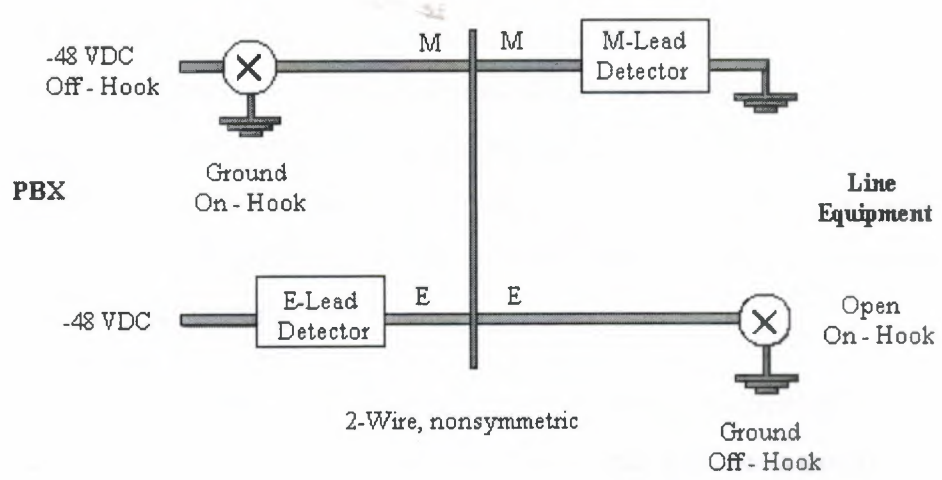

Figure 2.3 Type 1 E&M Signaling With a Two-Wire Line

Despite the simplicity of the in-band method, this type of signaling presented a number of problems. First, because the in-band signals by necessity fell within the bandwidth of speech signals, speech signals could at times interfere with the in-band signals. Second, in-band signaling did not always make efficient use of the available telephone circuits. For example, if a called party's telephone instrument was in use, the called party's central office would generate a busy signal that was carried by the already established voice path through the PSTN to the calling party's handset. Hence, a full voice-circuit path through the network was tied up merely to convey a busy signal.

2.4.2 Out–of–Band Signaling

Any transmission technology in which signaling is separate from the data being transmitted. Out-of-band signaling uses one or more channels for transmitting data or voice information and one special out-of-band channel for performing signaling functions such as establishing and terminating the communication link, controlling flow, or transmitting error information. The out-of-band channel can be:

- A physically separate set of wires (such as pins 4 and 5 of an RS-232 cable, which perform flow control functions and do not carry data)
- A multiplexed system in which bandwidth is divided into two or more channels within the same set of wires (such as Integrated Services Digital Network, in which the two B channels and one D channel are multiplexed onto the same set of wires)

The opposite of out-of-band is in-band, in which signaling information is sent over the same channel as the data transmission. Out-of-band transmission is usually considered a better choice than in-band transmission for the following reasons:

- None of the valuable data bandwidth is used for signaling.
- The data stream is not interrupted with signaling information.
- The signaling information cannot be disrupted by the noise created by the data transmission.
- Data transmission characters cannot accidentally (or purposefully) initiate control actions.

2.4.3 Single-Frequency and Multi-frequency Signaling

R1 and R2 signaling standards are used to transmit supervisory and address signaling information between voice network switches. They both use single-frequency signaling for transmission of supervisory information and multi-frequency signaling for addressing information [2].

2.4.3.1 R2 Signaling

R2 signaling specifications are contained in ITU-T Recommendations Q.400 through Q.490. The physical connection layer for R2 is usually an E1 (2.048 megabits per second [Mbps]) interface that conforms to ITU-T standard G.704. The E1 digital facilities carrier runs at 2.048 Mbps and has 32 time-slots. E1 time-slots are numbered TS0 to TS31, where TS1 through TS15 and TS17 through TS31 are used to carry voice, which is encoded with pulse code modulation (PCM), or to carry 64 kbps data. This interface uses time slot 0 for synchronization and framing (same as for Primary Rate Interface [PRI]) and uses time slot 16 for ABCD signaling. There is a 16-frame multi-frame structure that allows a single 8-bit time slot to handle the line signaling for all 30 data channels.

- **R2 Call Control and Signaling**

Two types of signaling are involved: line signaling (supervisory signals) and inter-register signaling (call setup control signals). Line signaling involves supervisory information (on-hook and off-hook) and inter-register signaling deals with addressing.

R2 uses channel-associated signaling (CAS). This means that, in the case of E1, one of the time slots (channels) is dedicated to signaling as opposed to the signaling used for T1. The latter uses the top bit of every time slot in every sixth frame.

This signaling is out-of-band signaling and uses ABCD bits in a similar manner to T1 robbed-bit signaling to indicate on-hook or off-hook status. These ABCD bits appear in time slot 16 in each of the 16 frames that make up a multi-frame. Of these four bits, sometimes known as signaling channels, only two (A and B) are actually used in R2 signaling; the other two are spare. In contrast to robbed-bit signaling types such as wink start, these two bits have different meanings in the forward and backward directions. However, there are no variants on the basic signaling protocol.

The transfer of call information (called and calling numbers, and so on) is performed with tones in the time slot used for the call (called in-band signaling).

R2 uses six signaling frequencies in the forward direction (from the initiator of the call) and a different six frequencies in the backward direction (from the party who answers the call). These inter-register signals are of the multi-frequency type with a two-out-of-six in-band code. Variations on R2 signaling that use only five of the six frequencies are known as decadic CAS systems.

Inter-register signaling is generally performed end-to-end by a compelled procedure. This means that tones in one direction are acknowledged by a tone in the other direction. This type of signaling is known as multi-frequency compelled (MFC) signaling.

There are three types of inter-register signaling [5]:

1. R2-Compelled - When a tone-pair is sent from the switch (forward signal), the tones stay on until the remote end responds (sends an ACK) with a pair of tones that signals the switch to turn off the tones. The tones are compelled to stay on until turned off.
2. R2-Non-Compelled - The tone-pairs are sent (forward signal) as pulses, so they stay on for a short duration. Responses (backward signals) to the switch (Group B) are sent as pulses. There is no Group A signals in non-compelled inter-register signaling.
3. R2-Semi-Compelled—Forward tone-pairs are sent as compelled. Responses (backward signals) to the switch are sent as pulses. This scenario is the same as compelled, except that the backward signals are pulsed instead of continuous.

Features that can be signaled include:

- Called or calling party number
- Call type (transit, maintenance, and so on)
- Echo-suppressor signals
- Calling party category
- Status

2.4.3.2 R1 Signaling

R1 signaling specifications are contained in ITU-T Recommendations Q.310 through Q.331. This document contains a summary of the main points. The physical connection layer for R1 is usually a T1 (1.544-Mbps) interface that conforms to ITU-T standard G.704. This standard uses the 193rd bit of the frame for synchronization and framing (same as T1).

- **R1 Call Control and Signaling**

Again two types of signaling are involved: line signaling and register signaling. Line signaling involves supervisory information (on-hook and off-hook) and register signaling deals with addressing.

R1 uses in-slot CAS by bit robbing the eighth bit of each channel every sixth frame. This type of signaling uses ABCD bits in an identical manner to T1 robbed-bit signaling to indicate on-hook or off-hook status.

The transfer of call information (called and calling numbers, and so on) is performed with tones in the time slot used for the call. This type of signaling is also called in-band signaling.

R1 uses six signaling frequencies that are 700 to 1700 Hz in 200-Hz steps. These inter-register signals are of the multi-frequency type and use a two-out-of-six in-band code. The address information contained in the register signaling is preceded by a KP tone (start-of-pulsing signal) and terminated by a ST Tone (end-of-pulsing signal).

2.4.4 Loop – Start Signaling

Loop-start signaling is a supervisory signaling technique that provides a way to indicate on-hook and off-hook conditions in a voice network. Loop-start signaling is used primarily when the telephone set is connected to a switch. This signaling technique can be used in any of these connections:

- Telephone set to CO switch
- Telephone set to PBX switch
- Telephone set to foreign exchange station (FXS) module (interface)
- PBX switch to CO switch
- PBX switch to FXS module (interface)
- PBX switch to foreign exchange office (FXO) module (interface)
- FXS module to FXO module

2.4.5 Ground – Start Signaling

Ground-start signaling is another supervisory signaling technique, like loop-start, that provides a way to indicate on-hook and off-hook conditions in a voice network. Ground-start signaling is used primarily in switch-to-switch connections. The main difference between ground-start and loop-start signaling is that ground-start requires ground detection to occur in both ends of a connection before the tip and ring loop can be closed.

*Although loop-start signaling works when you use your telephone at home, ground-*start signaling is preferable when there are high-volume trunks involved at telephone switching centers. Because ground-start signaling uses a request and/or confirm switch at both ends of the interface, it is preferable over FXOs and other signaling methods on high-usage trunks.

2.4.6 Common Channel Signaling (CCS)

In order to overcome the problems of other signaling methods and to speed the call set-up process in long-distance calls, another form of interoffice signaling, known as common channel signaling (CCS), was developed. The first version of CCS was developed between 1964 and 1968 by the International Telegraph and Telephone Consultative Committee (CCITT), a United Nations body that establishes worldwide telecommunications standards [2].

CCS is a signaling method in which a single channel conveys, by means of labeled messages, signaling information relating to a multiplicity of circuits, or other information, such as that used for network management. This migration to out-of-band signaling is only the beginning of a major network evolution with regard to signaling, which will culminate in the deployment of ISDN and BISD

- **CCS Architecture**

CCS involves a separate high-speed network ("the common channel") to transfer supervisory signaling information in an out-of-band fashion; this separate network carries signaling information from a layer number of different users, in multiplexed fashion, employing packet-switching technology. The separation of the signaling from the information channel, as well as the greater repertoire of command message formats, allows a more methodic migration of the network to any advanced architectural configuration; this follows from the fact that change can be made without the high cost associated with physical replacement or modification of hardware.

Because supervisory instructions are coded as messages, instead of some sequence of tones, and because of the higher bandwidth available for signaling, more detailed information about a call, in terms of desired network treatment, call origin, etc, can be exchanged across the network. In tern, this implies more sophisticated services. The talk-off problem is totally eliminated with the separate signaling facilities. Another advantage of CCS is that signals can be sent in both directions simultaneously, and during the conversation, if necessary. This last feature is very valuable for some advanced services. Business-case analyses also proved-in CCS on the merits of saving network equipment and trunks with faster signaling.

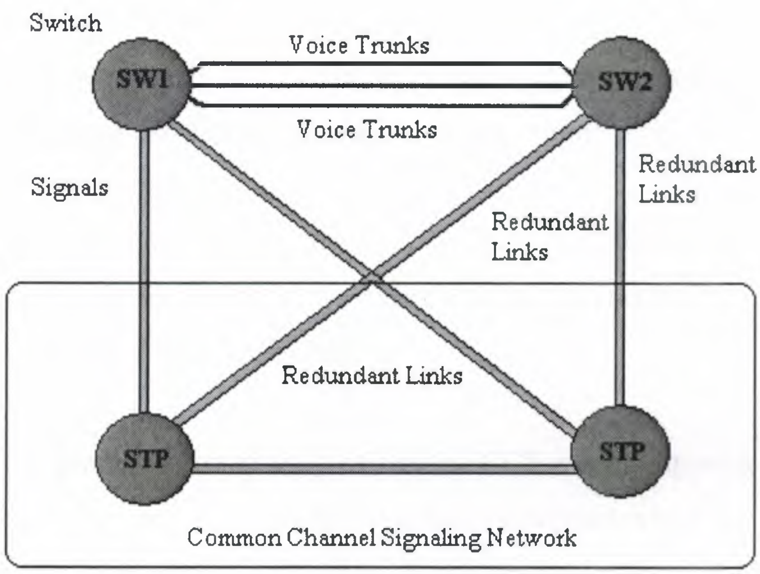

Figure 2.4 Links to the CCS Network

CCS allows access to many points in the network, not just switches. Advanced voice and data services may depend on remote data bases, processors, and facilities. Thus CCS provides:

1. Direct local office-to-local office signaling connectivity
2. Local office-to service node signaling connectivity

CCS is reliable and fast. It replaces both SF and the MF signaling equipment and methods; addressing digits are converted to data messages (packets). All modern digital switches can be equipped with CCS capabilities. The CCS interface is an electronic device that interprets incoming data messages and transfers the translated signal to the common control-call processor. Three signaling modes need to be considered:

1. Associated mode: in this mode, the messages relating to a particular signaling relation between two adjacent points are conveyed over a link directly interconnecting these signaling points.
2. Nonassociated mode: in this mode, the messages relating a particular signaling relation are conveyed over two or more links in tandem passing through one or more signaling points other than those which are the origin and the destination of the messages.
3. Quasiassociated mode: this mode is a subcase of the nonassociated mode. Here, the path taken by the messages through the network is predetermined and, at a given point in time, fixed.

A direct plant implementation of fully associated CCS would require a point-to-point signaling link between any two switches. Switches equipped with CCS interfaces can be interconnected with direct data links if the traffic volume of these signaling messages is high enough. Most switches, however, are connected to data communication packet switches. The nonassociated signaling is then implemented over a network that employs signal transfer points (STPs) operating as packet switches; this topology obviates the need for a larger number of point-to-point links or more STPs in tandem. Connectionless packet-switching techniques are employed. The function of the STPs is to route signaling messages between the various constituent links, without altering the message. Thus, the only functions are the level 2 error detection-correction task in signaling message content and the level 3 network routing function.

Until the advent of ISDN, out-of-band functional signaling will be limited to the trunk side of the plant. The SS7 signaling network only applies at the trunking level. With ISDNs D channels, a complementary (but not identical) capability will be extended to the end-user.

The STPs are packet switches that handle the routing of the signaling messages. As such, these nodes provide for concentration and ensuing efficiency: few switching offices in the inter-LATA or intra-LATA network have trunk groups large enough to justify direct connection of the signaling channels between the offices in question ("associated" or "direct connected" links). STPs may be redundant to ensure availability and reliability. While normally operating in a load-sharing mode, one STP can take over it the other files. In most cases, signaling messages are routed over one or two STPs.

2.5 CCITT International Signaling Systems

CCITT is an Abbreviation of (Consultative Committee on International Telephone & Telegraph), an organization that sets international communications standards. CCITT, now known as ITU (the parent organization) has defined many important standards for data communications [3]. Here are the most common international CCITT signaling systems used around the world from the early days of telephony until today:

1. **CCITT 1**

   An old international system, now deceased. Used a 500 Hz tone interrupted at 20 Hz (Ring) for 1-way line signals.

2. **CCITT 2**

   Proposed "International Standard" that never caught on much. Used 600 Hz interrupted by 750 Hz. Still used in Australia, New Zealand and South Africa

3. CCITT 3

An early in-band system that uses 2280 for both line and register. Used in France, Austria, Poland and Hungary.

4. CCITT 4

A variation of 3, but uses 2040 and 2400 for end to end Tx of line and register. Used for international Traffic in Europe, but cannot be used with TASI (AKA Multiplex or "that dammed clipping").

5. CCITT 5

This is the most popular, and the one used in the US. 2400 and the infamous 2600 are used for link to link (not merely end to end line signals. Registers are handled via DTMF (Touchtones).

6. CCITT 5 bis

Just like above, but an 1850 Hz tone is used for TASI locking and transmission of line signals.

7. CCITT 6

It uses digital data sent out-of-band to control the connection. In other words, the connection is made and billing started before you can get control. This system had many failures and problems in general.

8. CCITT 7

CCITT's Signaling System Number 7 (CCITT 7) is a common channel signaling system developed by ITU-T (formerly CCITT) in response to a demand for more features and integrated data services. It is a high-speed, out-of-band signaling system based on ITU-T recommendation Q.700 series that has become a global standard for telecommunications. SS7 defines the architecture, procedures, and protocols for information exchange over digital channels. It is designed to support call setups, routing, billing, database information, and special service functions for PSTNs. The ITU-T definition of SS7 allows for national variants such as ANSI, Bellcore (North America), ETSI (used in Europe), and several country-dependant variants.

Examples of some applications supported by SS7 are:

- PSTN
- ISDN (Voice and Data)
- Interaction with network databases and service control points for service control
- Mobile services
- Operations administration and maintenance of networks

2.6 Summary

In this chapter we described what signaling is and the need of signaling in the telecommunication networks. We discussed different signaling techniques which were used in old telephone systems up to the most effective methods used now days. And finally we mentioned the signaling system #7 which will be the topic of our next chapter.

3. SIGNALING SYSTEM NUMBER 7

3.1 Overview

The SS#7 protocols were developed by AT&T since 1975 and defined as standard by ITU-T during 1981 in ITU-T's Q.7XX-series recommendations. SS#7 is the protocol used by the telephone companies for interoffice signaling. In the past, in-band signaling techniques were used on interoffice trunks. This method of signaling used the same physical path for both the call-control signaling and the actual connected call. This method of signaling is inefficient and is rapidly being replaced by out-of-band or common-channel signaling techniques.

A network utilizing common-channel signaling is actually two networks in one:

1. First there is the circuit-switched "user" network which actually carries the user voice and data traffic. It provides a physical path between the source and destination.
2. The second is the signaling network which carries the call control traffic. It is a packet-switched network using a common channel switching protocol.

The original common channel interoffice signaling protocols were based on Signaling System Number 6 (SS#6). Today SS#7 is being used in new installations worldwide. SS#7 is the defined interoffice signaling protocol for ISDN. It is also in common use today outside of the ISDN environment.

The primary function of SS#7 is to provide call control, remote network management, and maintenance capabilities for the inter-office telephone network. SS#7 performs these functions by exchanging control messages between SS#7 telephone exchanges (signaling points or SPs) and SS#7 signaling transfer points (STPs).

The switching offices (SPs) handle the SS#7 control networks as well as the user circuit-switched network. Basically, the SS#7 control network tells the switching office which paths to establish over the circuit-switched network [6].

3.2 SS7 Network Architecture

While the PTSN has a number of key elements, it really is the switching location that makes it a network. Switches are the "glue" that holds the PSTN together. The SS7 signaling architecture consists of three essential components, interconnected via signaling links [7].

3.2.1 SS7 Nodes

There are three main types of nodes connecting the SS7 network. These are STP, SCP and SSP.

3.2.1.1 The STP (Signaling Transfer Point)

STPs are packet switches, and act like routers in the SS7 network. Messages are not usually originated by an STP. An STP can act like a firewall, screening messages with other networks.

There is no need for connection in the SS7 network. What is referred to as "circuits" in the PSTN can not carry messages until the switch makes a physical connection. Instead of circuits, the SS7 makes use of transmission lines called links. In concept, at least, these links always exist and are always available to carry messages. Instead of "connecting," the STP needs only to direct messages to the links which it selects as most appropriate to deliver the message.

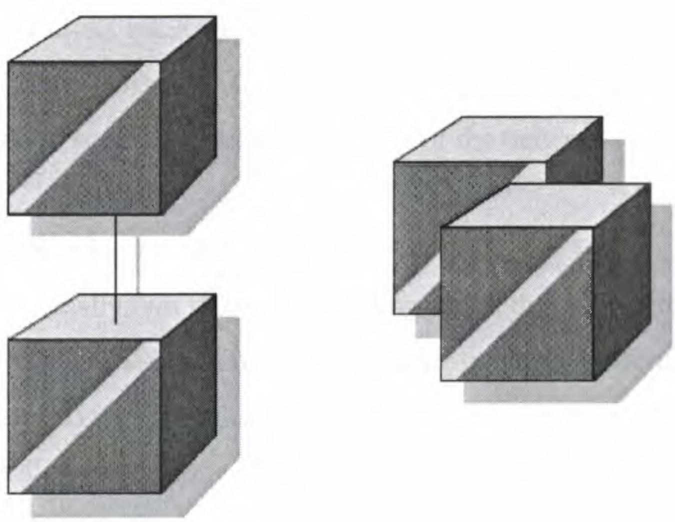

Figure 3.1 Graphic Representations of STPs

STPs route SS7 messages (based on information contained in the message format) to outgoing signaling links over the SS7 network. They are the most versatile of all the SS7 entities, and are a major component in the network.

There are three levels of STPs [7]:

1. National Signal Transfer Point: A National STP exists within the national network (will vary with the country). It can transfer messages that use the same national standard of protocol. Messages can be passed to an International STP, but can not be converted by the National STP. Protocol converters often interconnect a National and an International STP by converting from ANSI to ITU-TS.

2. International Signal Transfer Point: An International STP functions within an international network. It provides for SS7 interconnection of all countries, using the ITU-TS standard protocol. All nodes connecting to an International STP must use the ITU-TS protocol standard.

3. Gateway Signal Transfer Point: A Gateway STP converts signaling data from one protocol to another. Gateway STPs are often used as an access point to the international network. National protocols are converted to the ITU-TS protocol standard. Depending on its location, the Gateway STP must be able to use both the International and National protocol standards. A Gateway STP also serves as an interface into another network's databases, such as from an inter-exchange carrier (IXC) to an end office. The Gateway STP can also be configured to screen for authorized users of the network.

3.2.1.2 The SSP (Service Switching Point)

There are actually two types of Signaling nodes that are switch associated. The first type is called a CCSSO (Common Channel Signaling Switching Office). These are end or tandem offices which have the capability to use the SS7 in what is referred to as a trunk signaling mode for call set-up. The second type (and the name we'll hear most often) is the Service Switching Point (SSP).

SSPs create packets (signal units) and send those messages to other SSPs, as well as queries to remote shared databases to find out how to route calls. They can originate, terminate, or switch calls. SSPs communicate with the voice switch via the use of primitives and have the ability to send messages using ISUP (call setup and teardown) and TCAP (database lookup) protocols.

The SSP uses the calling party information (dialed digits) to determine how to route the call. It looks up the dialed digits in the SSP routing table to find the corresponding trunk circuit and terminating exchange. The SSP then sends an SS7 message out to the adjacent exchange requesting a circuit connection on the trunk which was specified in the routing table.

The adjacent exchange sends an acknowledgement back, giving permission to use that trunk. Using the calling party information contained in the setup info, the adjacent exchange determines how to connect to the final destination. This might require several trunks to be set up between several different exchanges. SSP manages all of these connections until the destination is reached. Figure 3.2 shows the SSP connection in the network.

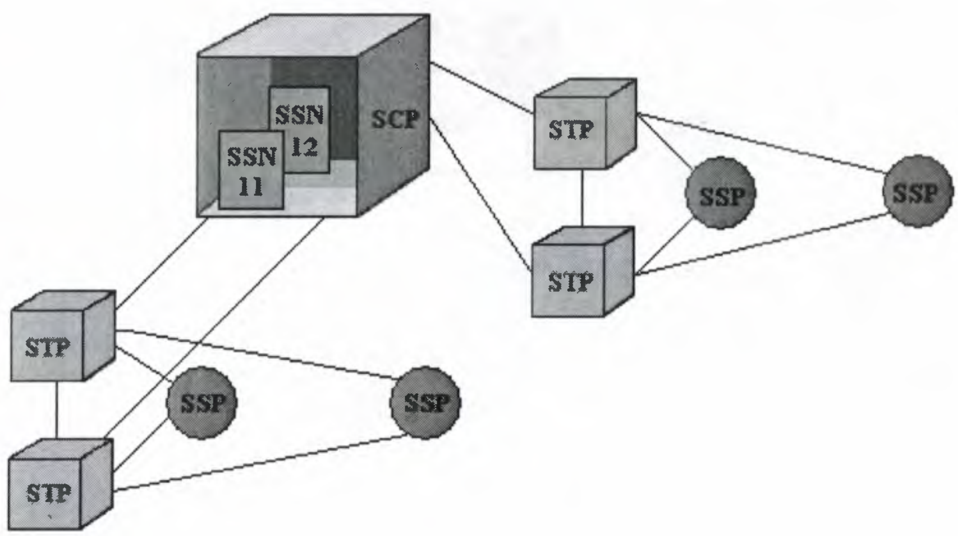

Figure 3.2 SSP Connections in the SS7 network

3.2.1.3 The SCP (Service Control Point)

An SCP is usually a computer used as a front end to a database system. In today's network you will find a database wherever a translation, verification, or simply information is required. At the doorway to that database you will find a Service Control Point. This is the node that provides the mechanisms for data to be retrieved from the database in a form that is suitable to the purposes of the node initiating the query. Since the types of services that can be offered are limited only by imagination and available data, it is likely that SCPs will continue to play a significant role in the growth of the SS7 Network.

In general, the major databases (like the 800 database) have been centralized in the network. That is not to say that a single such database exists; but, rather, that several identical databases exist throughout the network. Obviously each of these databases should contain the same information. The address of an SCP is a point code, and the address of the database it interfaces with is a subsystem number. The database is an application entity which is accessed via the TCAP protocol.

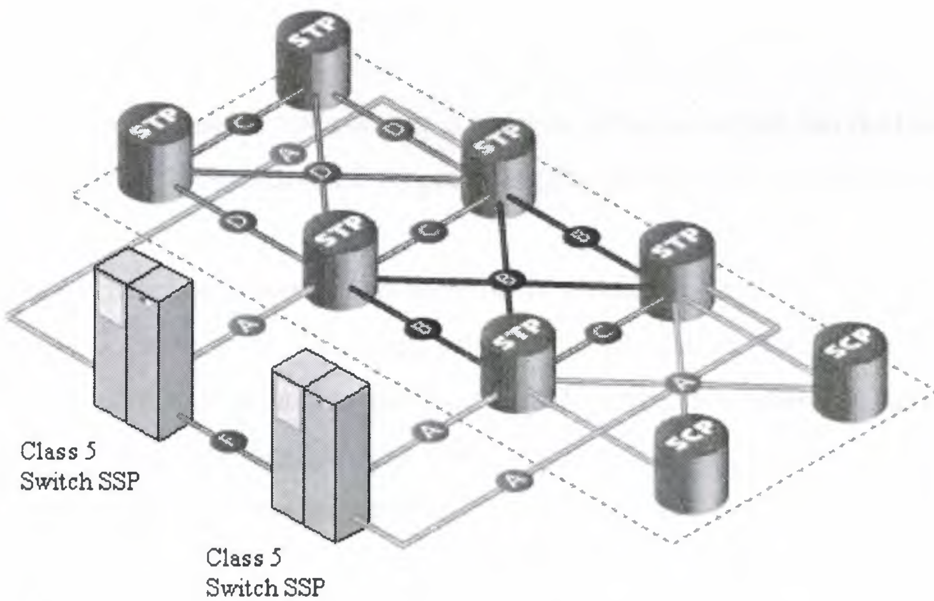

Figure 3.3 the SS7 Network Backbone with SCP Nodes

- **The CRP (Customer Routing Point)**

  When a switch makes a query to obtain switch routing information based on the dialed digits of 800 numbers assigned to this company, the query is routed to the database the company maintains for itself. Since this phone customer establishes a signaling point for the purpose of providing its own routing information, the node is called a Customer Routing Point (CRP).

The company operating the CRP has full control of the routing information being returned to the switch. When it becomes necessary to change the routing (as in the case of the burned out location), the company simply updates its database. The minute the company becomes aware of an inaccessible location, the changes can be made, and the very next call will be answered at a new location.

- **The IP (Intelligent Peripheral)**

  The same second level addressing capability allows the SCP to access and make available services located at other signaling points in the network. Sometimes this entails invoking features for which the switch is not equipped. At other times it entails utilizing an Intelligent Peripheral.

In general, the Intelligent Peripheral (IP) is home to a Process which can deal with the requests made of it through the SCP by providing the services of a variety of devices.

Another node which is important to mention is called the Services Node (SN). In some networks there is no difference between an IP and an SN. However, it is generally agreed that what makes the node an SN is the programmable services it offers rather than the physical devices. Still, what one network calls an IP might be called a Services Node in another network.

- **The MSC (Mobile Switching Center)**

  Mobile Networks normally end up with numerous nodes in SS7 networks. The Mobile Switching Center communicates with and controls the radio transceivers which form the cells of a cellular network. Usually, once the transceiver has received and sent calls to the cell phone, the wireless part of a wireless network has done all it can do. The next step is for the MSC to make a circuit connection into the PSTN for an outgoing call or to accept a connection from the PSTN for an incoming call.

To provide the customer information required for other networks to validate a call, and to keep subscriber data necessary for the local network to provide numerous services, another node called the Home Location Register (HLR) is deployed. This node is essentially a database providing subscriber information.

Mobile networks employ other SS7 nodes as well. Authentication Centers (AUC) provide security processes to verify and validate cell phones seeking services. Short Message Centers (SMC) communicate with HLRs and MSCs to coordinate delivery of the text messages they store. All of these make use of the SS7 to send the messages they need to send to each other.

Table 3.1 Telco Databases Accessible via SCP

Abbreviation	Name	Description
BSDB	Business Services Database	Allows companies to create and store proprietary databases, as well as create private networks
CMSDB	Call Management Services Database	Provides information relating to call Processing, network management (prevent congestion), call sampling (create reports for traffic studies), and the routing, billing and third-party billing for 800, 976 and 900 numbers.
VLR	Visitor Location Register	Used when a cell phone is not recognized by the mobile switching center (MSC).
LIDB	Line Information Database	Provides billing instructions.
LNP	Local Number Portability	Allows people to change Telco service Providers but keep their same telephone number.
OSS	Operations Support Systems	Associated with remote maintenance centers for monitoring and managing SS7 and voice networks.

3.2.2 SS7 Links

An SS7 link is the physical transmission line (serial 56/64 Kbps or DS0 channel) that connects the individual nodes in an SS7 network [7].

SS7 networks are built to be highly reliable and redundant. Link diversity is built into the network design, providing multiple signaling paths, so that there is no single point of failure. This practice ensures that redundant links have the capacity to handle all rerouted network traffic.

3.2.2.1 A-Links

Access links (A-links) interconnect an STP and either an SSP or an SCP (signaling end points). Their sole purpose is to deliver signaling to and from signaling end points. End points always have at least two A-links (also called signaling beginning points). Any signaling that an SSP or SCP needs to send to any other node in the SS7 network is sent on one of its A-links to its "home" STP, which processes and routes the message along its way. Messages addressed to an SSP or SCP are routed to its "home" STP, which forwards them to the addressed node over its A-links.

3.2.2.2 B- and D-Links

Bridge links (B-links) are the quad of links interconnecting peer pairs of STPs. Diagonal links (D-links) are the quad of links interconnecting mated pairs of STPs at different hierarchical levels. Since the SS7 network has no clear hierarchy, these links are referred to as B-links, D-links, or B/D-links.

3.2.2.3 C-Links

Cross links (C-links) interconnect mated STPs and are used to enhance the reliability of the signaling network not regularly used by SS7 traffic. They are used only when there has been a link failure which causes an STP to have no other route.

3.2.2.4 E- and F-Links

Extended links (E-links) connect an SSP to an alternate STP to provide backup connectivity to the network if the SSP's "home" STP cannot be reached on its A-link. Fully associated links (F-links) directly connect two signaling end points (SSPs and/or SCPs). They are not usually used in networks with STPs because they allow associated signaling only, thus bypassing the security features provided with an STP.

3.2.2.5 Linksets

Links are put into groups called linksets. Up to 16 links can be assigned to one linkset. All links in a linkset must have the same adjacent node. (See Figure 3.4) Switches will alternate traffic across all links in a linkset to ensure equal usage of all facilities in the network

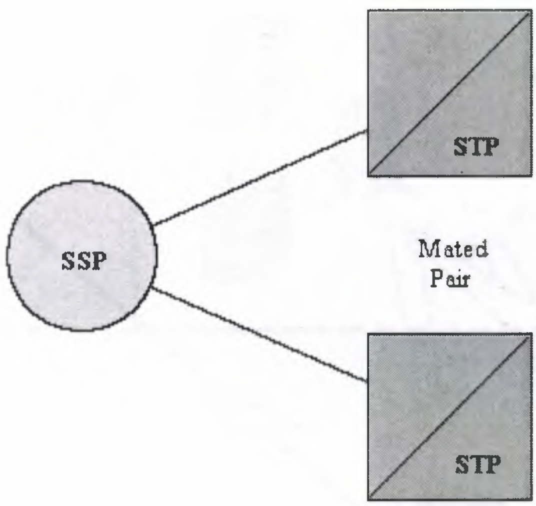

Figure 3.4 Linksets

- **One Final Look**

With our discussion of SS7 network architecture complete, all that remains is to illustrate the network in all of its major aspects and to review the nodes which have been discussed. Figure 3.5 illustrates the SS7 network and its interface to the PSTN and Wireless networks.

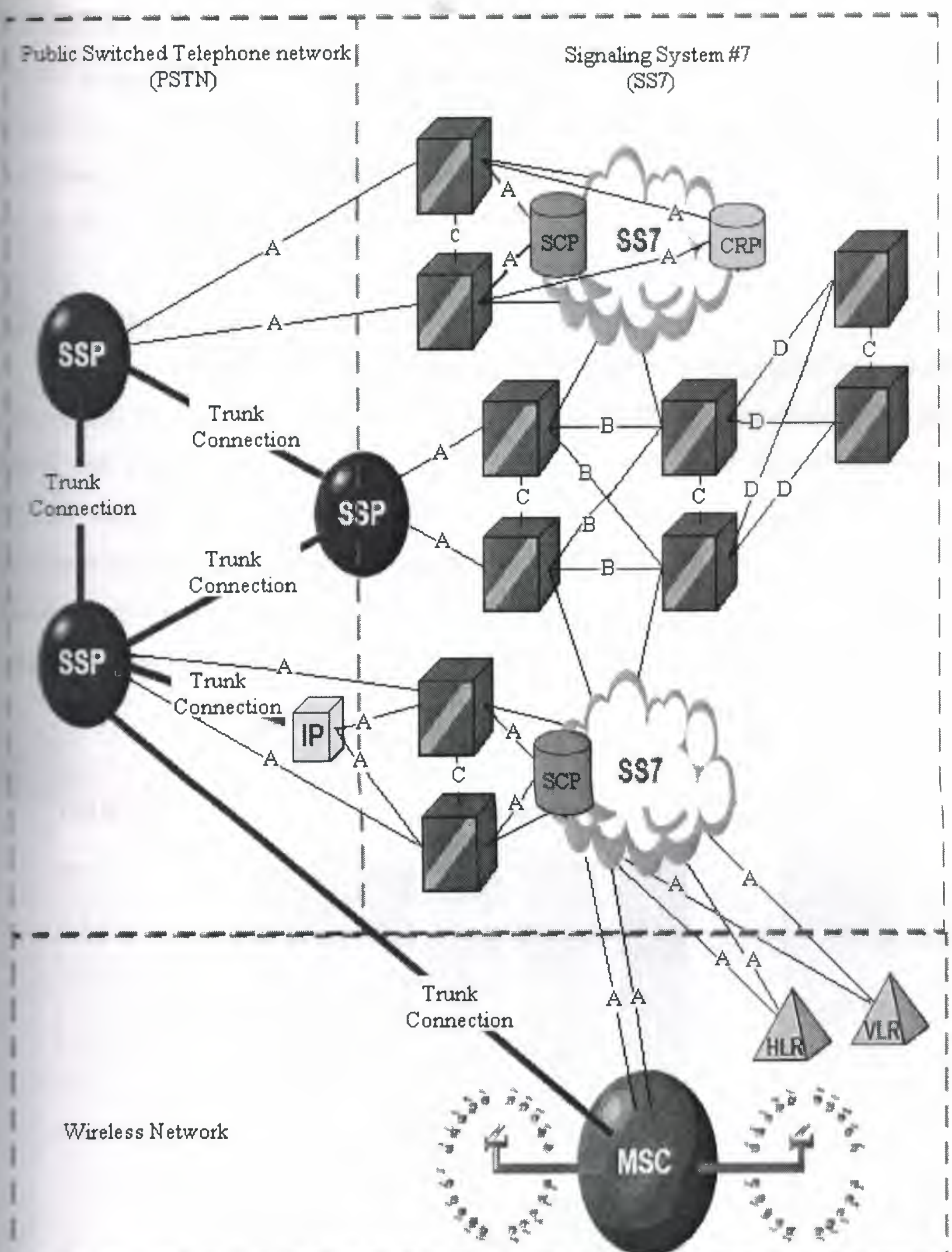

Figure 3.5 SS7 network and its interface to the PSTN and Wireless networks

3.3 SS7 Protocol Stack

A stack is a set of data storage locations that are accessed in a fixed sequence. The SS7 stack is compared against the Open Systems Interconnection (OSI) model for communication between different systems made by different vendors. Figure 3.6 shows the components of the SS7 protocol stack.

3.3.1 SS7 Physical Layer (Level 1)

This is the physical level of connectivity, virtually the same as Layer 1 of the OSI model. SS7 specifies what interfaces will be used, both Bellcore (Telecordia) and ANSI call for either the DS0A or the V.35 interface. Because central offices are already using DS1 and DS3 facilities to link one another, the DS0A interface is readily available in all central offices, and is preferred in the SS7 network. As the demands on the SS7 network increase (local number portability), and as the industry migrates toward ATM networks, the DS1 interface will become the link interface.

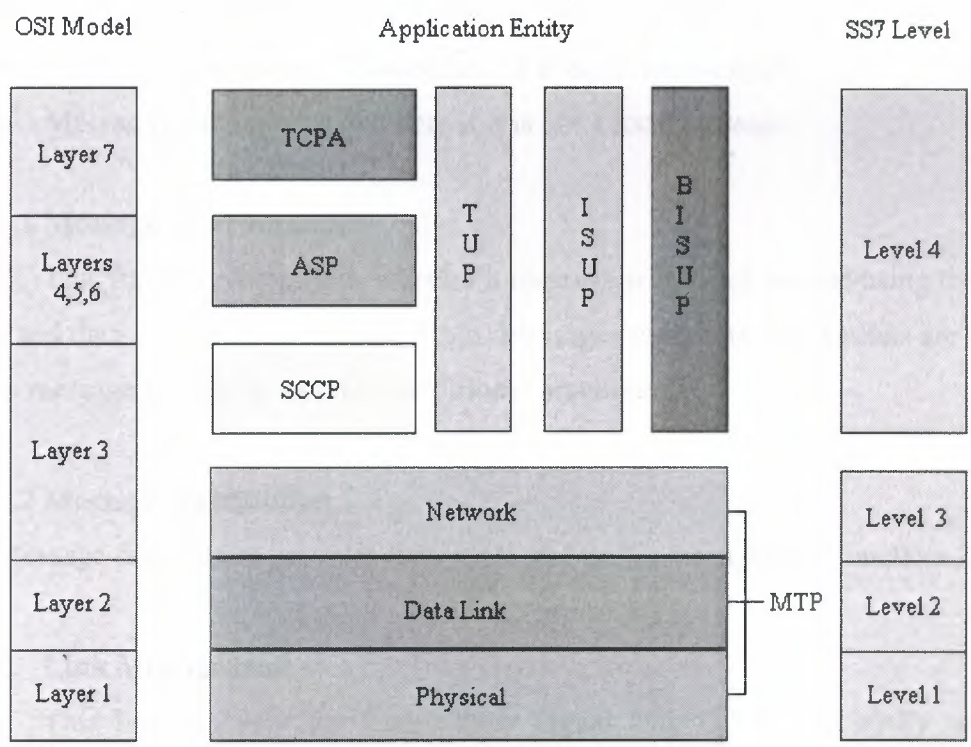

Figure 3.6 SS7 Protocol Stack

3.3.2 SS7 Data Link (Level 2)

The data link level provides the network with sequenced delivery of all SS7 message packets. Like the OSI data link layer, it is only concerned with the transmission of data from one node to the next, not to its final destination in the network [9].

Sequential numbering is used to determine if any messages have been lost during transmission. Each link uses its own message numbering series independent of other links. SS7 uses CRC-16 error checking of data and requests retransmission of lost or corrupted messages. Length indicators allow Level 2 to determine what type of signal unit it is receiving, and how to process it.

3.3.3 SS7 Network Layer (Level 3)

The network level depends on the services of Level 2 to provide routing, message discrimination and message distribution functions [9].

- Message Discrimination determines to whom the message is addressed.
- Message Distribution is passed here if it is a local message.
- Message Routing is passed here if it is not a local message.

3.3.3.1 Message Discrimination

This function determines whether a message is local or remote using the point code and data contained in a lookup table. Messages to remote destinations are passed to the message routing function for additional processing.

3.3.3.2 Message Distribution

Message distribution provides link, route and traffic management functions.

1. **Link Management**

This function uses the Link Status Signal Unit (LSSU) to notify adjacent nodes of link problems. Level 3 will send LSSUs via Level 2 to the adjacent node, notifying it of the problems with the link and its status. Diagnostics consists of realigning and resynchronizing the link.

- **Realignment**: All traffic is removed from the link, counters are reset to zero, timers are reset and Fill-In Signal Units (FISUs) are sent in the meantime (called the proving period).
- **Proving Period**: Amount of time FISUs are sent during link realignment. The duration of the proving period depends on the type of link used. Bellcore specifies the proving period for a 56 Kbps DS0 link is 2.3 seconds for normal proving and 0.6 seconds for emergency proving.

Another form of link management uses changeover and change back messages sent using Message Signal Units (MSUs). MSUs advise the adjacent node to send traffic over another link within the same linkset. The alternate link must be within the same linkset. The bad link is being realigned by Level 3 while traffic is rerouted over alternate links. Change back message is sent to advise the adjacent node that it can use the newly restored link again. Change back messages are typically followed by a change back acknowledgement message.

2. **Route Management**

This function provides a means for rerouting traffic around failed or congested nodes. Route management is a function of Level 3 and works together with link management. Route management informs other nodes of the status of the affected node. It uses Message Signal Units (MSUs) generated by adjacent nodes and is not usually generated by the affected nodes. (Link management only informs adjacent nodes.)

3. **Traffic Management**

This function provides flow control if a node has become congested. It allows the network to control the flow of certain messages based on protocol. Traffic management deals with a specific user part within an affected node.

For example, if ISUP is not available at a particular node, a traffic management message can be sent to adjacent nodes informing them that ISUP is not available, without affecting TCAP messages on the same node.

3.3.3.3 Message Routing

Message discrimination in Level 3 will pass messages to message routing if it determines the message is not local. Message routing reads the called and calling party addresses to determine the physical address in the form of a point code. Every SS7 node must have its own unique point code. Message routing determines the point code from an address contained in the routing table.

3.3.4 Message Transfer Part

Protocols are used within the layers (levels) of the SS7 protocol to accomplish functions called for at each level. Levels 1, 2 and 3 are combined into one part, the Message Transfer Part (MTP). MTP provides the rest of the levels with node-to-node transmission, including basic error detection and correction schemes and message sequencing. It provides routing, message discrimination and distribution functions within a node. Figure 3.7 shows the message transfer part.

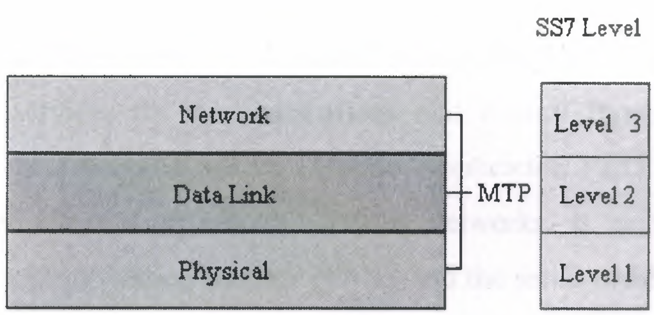

Figure 3.7 Message Transfer Part Components

3.3.5 SS7 Protocols, User and Application Parts (level 4)

Level 4 consists of several protocols, user parts and application parts. Figure 3.8 represents the level 4 parts. A brief explanation will be giver for each part of this level.

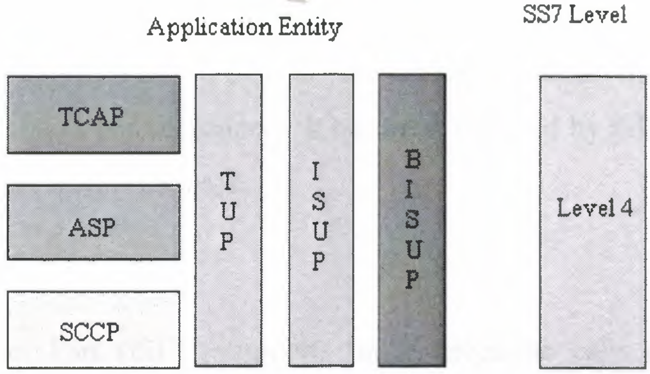

Figure 3.8 Level 4 Protocols, User & Application Part

3.3.5.1 TCAP

Transactional Capabilities Application Part (TCAP) facilitates connection to an external database. Information/data received is sent back in the form of a TCAP message. TCAP also supports remote control—ability to invoke features in another remote network switch.

OMAP (Operations, Maintenance and Administrative Part) is an applications entity that uses TCAP services for communications and control functions through the network via a remote terminal. MAP (Mobile Application Part) is used to share cellular subscriber information among different networks. It includes information such as the mobile identification number (MIN), and the serial number of the cellular handset. This information is used by the IS-41 protocol during cellular roaming.

3.3.5.2 ASP

Application Service Part (ASP) provides the functions of Layers 4 through 6 of the OSI model. These functions are not presently required in the SS7 network, and are under further study. However, the ITU-T and ANSI standards do reference ASP as viable.

3.3.5.3 SCCP

Signaling Connection Control Part (SCCP) is a higher level protocol than MTP that provides end-to-end routing. SCCP is required for routing TCAP messages to their proper database.

3.3.5.4 TUP

Telephone User Part (TUP) is an analog protocol that performs basic telephone call connects and disconnect. It has been replaced by ISUP, but is still used in some parts of the world (China).

3.3.5.5 ISUP

ISDN User Part (ISUP) supports basic telephone calls connect/disconnect between end offices. Used primarily in North America, ISUP was derived from TUP, but supports ISDN and intelligent networking functions. ISUP also links the cellular and PCS network to the PSTN. BISUP (Broadband ISUP) will gradually replace ISUP as ATM is deployed. Figure 3.9 illustrates the ISUP procedures.

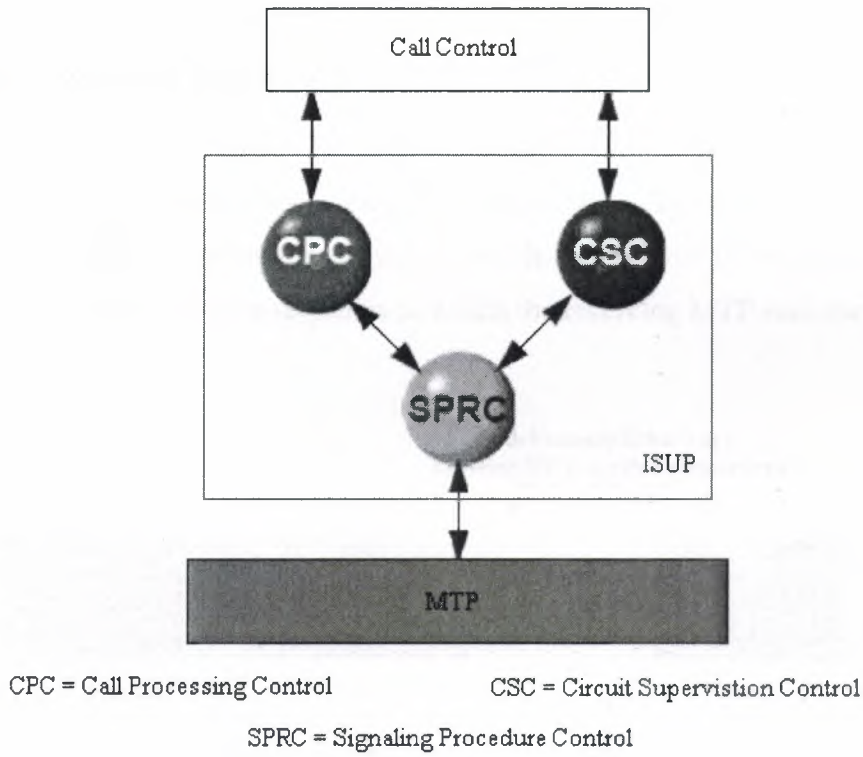

Figure 3.9 ISUP Procedure

3.3.5.6 BISUP

Broadband ISDN User Part (BISUP) is an ATM protocol intended to support services such as high-definition television (HDTV), multilingual TV, voice and image storage and retrieval, video conferencing, high-speed LANs and multimedia.

3.4 Signaling Units

The SS7 uses only three packets (signal units) in transmission. The majority of the fields are identical in each of these units. We'll examine each unit, and hopefully, end up with an understanding of the ways in which the protocol manages to deliver information from node to node and to the final destination of the message.

3.4.1 The Message Signal Unit (MSU)

We will start with the most complex of the signal units. While it may seem strange to work from the complex to the simple, I do so with a purpose. Most of the fields that appear in the Message Signal Unit appear in the other two signal units also. Once we understand the MSU, we'll need only some brief explanation to understand the other two as well.

3.4.1.1 Message Signal Unit Fields

We will examine these fields shown in figure 3.10 to determine the purpose of each and what kind of data belongs there. Notice the arrow indicating the direction of the data flow. This indicates the sequence in which the unit is assembled by the transmitting MTP and also the sequence in which the receiving MTP sees the data.

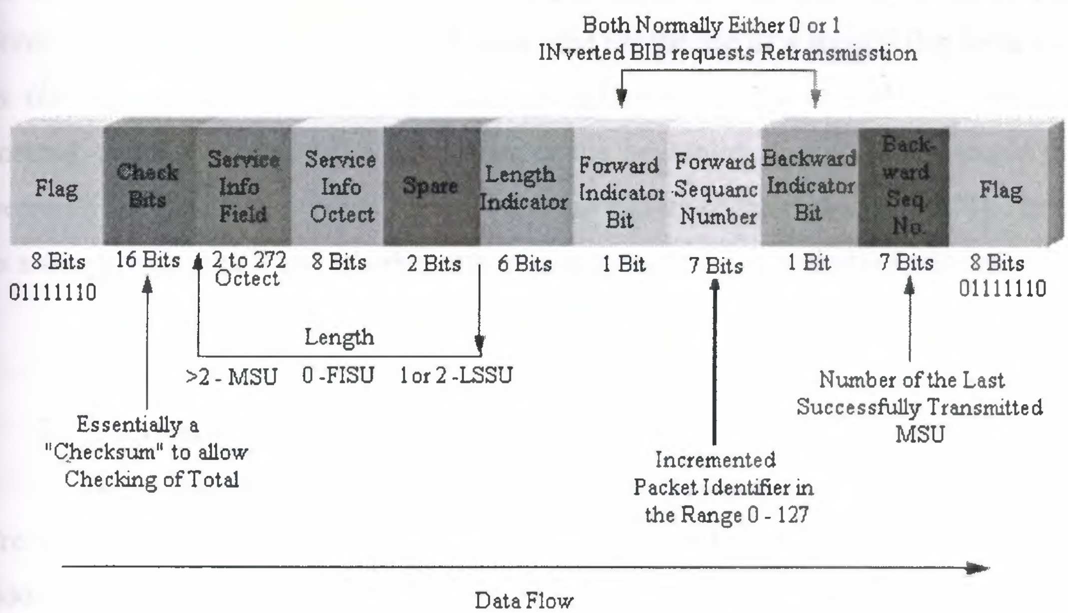

Figure 3.10 Message Signal Unit Field

1. The Flag

When dealing with a digital protocol, knowing where to start reading is of paramount importance. In the SS7 protocol the key to making the correct assumptions about the data lies in knowing exactly what data is being read. That, in turn, lies in knowing exactly where the data lies within the signal unit. For those reasons, there can be no confusion about where to start reading the message.

This is so important that the MTP can take no chances that it might start to read the data at some point other than the actual start of the signal unit. To avoid this, a unique eight bit code is placed at the beginning of the unit. This code is a byte with zeros at either end and six ones in the middle. There may be considerable data in the remainder of the signal unit. It is very likely that a zero followed by six ones and a zero will occur elsewhere in the signal unit. To ensure that there are no false flags, the sending MTP reads through the signal unit. Each time it reads five ones in a row, it inserts a zero. This is a procedure known as "bit stuffing." The MTP then attaches the flag and transmits the message.

At the receiving end, the MTP sees the flag and begins its reading of the signal unit. Every time it sees five ones, it removes the following zero. In this way, the confusion of multiple flags is eliminated, and the signal unit ends up restored to its original form. The original standards (CCITT) provided for the use of a second flag to be used at the end of the signal unit. The later ANSI standard saw no value in this, and instead, supports the use of a single flag at the beginning. In this way a single flag becomes both the beginning of one unit and the ending of the previous one. The result is a shorter signal unit and a higher rate (signal units per time period) of transmission.

2. Backward Sequence Number

From this point on, it will be helpful if you equate the term "backward" with "receiving node" and the term "forward" with "transmitting node". It is the receiving node which makes changes to this value (the BSN). It does so when it is returning a signal unit to positively acknowledge the receipt of a unit or to make a negative acknowledgment of a unit. In the latter case, the MTP will usually also request that the message be retransmitted. This will become clearer when we examine the Forward Sequence Number.

3. Backward Indicator Bit

Once again, it is the receiving node that will make changes to this value (the BIB). It will change this bit to the opposite of Forward Indicator Bit in the same signal unit being used to send a negative acknowledgment back to the transmitting side. The transmitting side reads this changed bit state as a request for retransmission.

4. Forward Sequence Number

This time it is the transmitting node which makes changes to the value. The transmitting MTP maintains a numbering resource which provides cyclical and sequential values in the range of zero to one hundred twenty seven (0 - 127). It places the value into this field and then simultaneously transmits the signal unit and copies it into a retransmit buffer. This provides the receiving side with a value by which to refer to the signal unit.

5. Forward Indicator Bit

Once again, it is the transmitting node that deals with this value. On transmission it ensures that the Forward Indicator Bit matches the Backward Indicator Bit.

6. Error Correction

The standards support two methods of correcting errors. In one (the Preventive Cyclic Retransmission Error Correction Method) the messages are retained on the transmitting side until acknowledged. During every break in transmission (no messages to be sent) the transmitting side simply retransmits all messages that have not yet been acknowledged. This method is generally used only for satellite transmission.

The other method is the Basic Error Correction Method. Now that we have seen the first five Message Signal Unit (MSU) fields, we'll follow this method through its sequence to see what both the transmitting MTP and the receiving MTP need to do to ensure the delivery of good messages.

First the transmitting side uses its numbering resource to provide a value for the Forward Sequence Number. Then it transmits the MSU and sends a copy to its retransmit buffer. The receiving MTP, of course, monitors the incoming message. If no error occurs, it will send an acknowledgment before it has seen the entire series of 128 signal units (0 - 127) applied by the transmitting side. This needs to be done because the transmitting side will not apply a value to a new message if that number matches a value in its retransmit buffer. If this occurs, the transmit node simply stops transmitting and the MTP indicates a "link failure."

For the **acknowledgment**, the MTP uses whichever signal unit it would normally be returning to the transmitting node. Since the MTP normally reports link status periodically, this would commonly be a Link Status Signal Unit (LSSU). To make the acknowledgment, the receiving MTP takes the Forward Sequence Number for the last valid signal unit and copies it to the Backward Sequence Number field of the signal unit it is returning. It leaves the Backward Indicator Bit alone so that the Forward and Backward Indicator Bits are returned as received (both the same).

When the signal unit arrives at the transmitting side, it is recognized as an acknowledgment without request for retransmission. The transmitting side now simply deletes from its retransmit buffer all signal units having the value of the Backward Sequence Number and all prior Sequence Numbers. When the receiving MTP detects an error, it once again uses the next planned return signal unit and copies the Forward Sequence Number of the last valid signal unit into the Backward Sequence number. This time, it toggles (from 1 to 0 or from 0 to 1) the Backward Indicator Bit so that it is no longer the same as the Forward Indicator Bit. When this unit is received at the transmitting node, it recognizes the unequal Forward and Backward Indicator Bits as a request for retransmission. It deletes all signal units with a value equal to or less than that of the Backward Sequence Number and begins retransmission of all signal units beginning with the one that is one higher than that of the Backward Sequence Number. Transmission is halted until the retransmission is complete.

7. Length Indicator

This field may seem a little strange. For one thing, it doesn't do what it was originally intended to do. The original intention was for this field to indicate the number of octets of significant data which followed it. That data was located in the Service Information Octet and the Signaling Information Field.

Signaling System #7 allows up to 272 octets of data in the Signaling Information Field alone. The Signaling Information Field is where the significant data that represents the actual message is placed. Of what use is a length indicator that can only represent values from 0 to 63? The answer is that as a length indicator it is of little use, but as an identifier of Signal Unit type it is very helpful. Signaling System #7 uses three types of signal units. The least complex signal unit is the Fill in Signal Unit (FISU) which contains no message or service data beyond the Length Indicator. Therefore, a length indication value of 0 identifies the Fill in Signal Unit.

Another, slightly more complex signal unit is the Link Status Signal Unit (LSSU). It has a Link Status Field which can contain one or two octets. Therefore, a length indication value of 1 or 2 identifies the Fill in Signal Unit.

Finally, the Message Signal Unit, as we have said, may contain much more data in the fields following the Length Indicator than can be counted using six bits. In that case, the counter simply goes as far as it can. If more than 64 octets are found, the value stays at 63 (0 to 63 range). The Signaling Information Field will never have less than 2 octets of data. Taken with the Service Information Octet, that means that the length indication will always be more than 2 if the signal unit is an MSU. The meaning then, of the Length Indicator value is as follows:

Length Indicator = **0 Fill In Signal Unit**

Length Indicator = **1** or **2 Link Status Signal Unit**

Length Indicator > **2 Message Signal Unit**

Any missing data is easily assessed using the sophisticated algorithm found in the Check Bits, so there is now no loss resulting from the fact that the Length Indicator cannot accurately indicate the amount of data in message fields.

8. Spare

From figure 3.10 we can see the spare is simply a two bit field. Its only purpose, generally, is to keep the entire signal unit to an even number of octets. Each field which is not an octet (FSN, BSN) has its counterpart which brings the number to eight (FIB, BIB). The spare becomes the counterpart for the six bit Length Indicator. This is particularly helpful at specific times. For example, when the MTP is aligning (or restoring) a link, it sends Fill In Signal Units and looks for errors. With no significant data, the total data must be evenly divisible by eight. A single bit (or 2, 3,4,5,6 or 7) lost or gained will change this no-remainder division and the unit can be adjudged invalid.

9. Signaling Information Octet

The Signaling Information Octet is apportioned into sub-fields of four bits, two bits and two bits. Figure 3.11 represents this field. We will examine each of the sub-fields in greater detail.

- **Service ID** - The Service ID field provides information about the type of message being sent. This provides level 3 with the necessary information for message distribution. Codes are provided for the user parts with the exception of TCAP. The reason for this is that TCAP messages are always appended to SCCP messages. An SCCP code along with a UNITDATA message type indicates a TCAP message. Other codes are provided when the message is a signaling network message dealing with management messages or with testing and maintenance messages.

- **Sub-Service Field -** This field contains bits for the indication of network in which the message sender is deployed. It also provides a spare for network dependent usage.

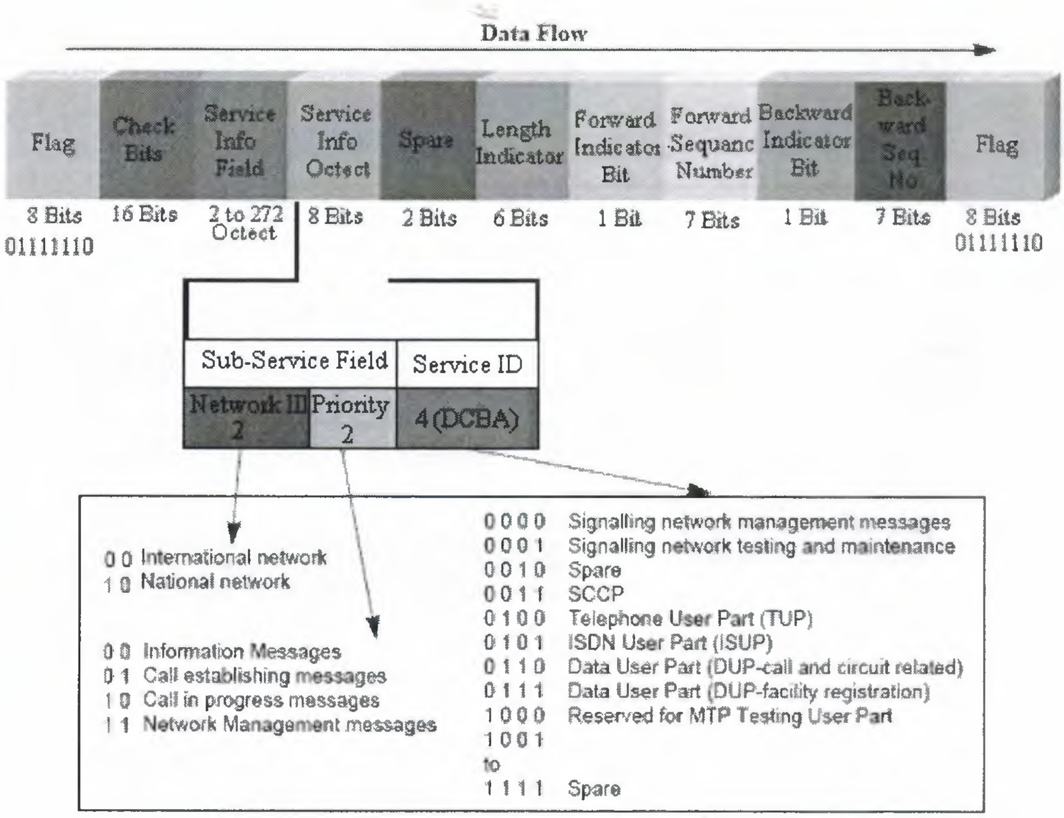

Figure 3.11 Signaling Information Octet Field

10. Signaling Information

Every field that we have looked at so far is designed to control the delivery and provide further specific information about the data contained in this field. This is the message field of the Message Signal Unit. In the U.S. this field never contains less than 2 octets of data, nor does it ever contain more than 272 octets. This allows the transmission of 265 octets of information along with a label. There may also be additional housekeeping information. Such information may be used at level 4 to link information blocks together (among other possible uses).Figure 3.12 will begin our examination of this field.

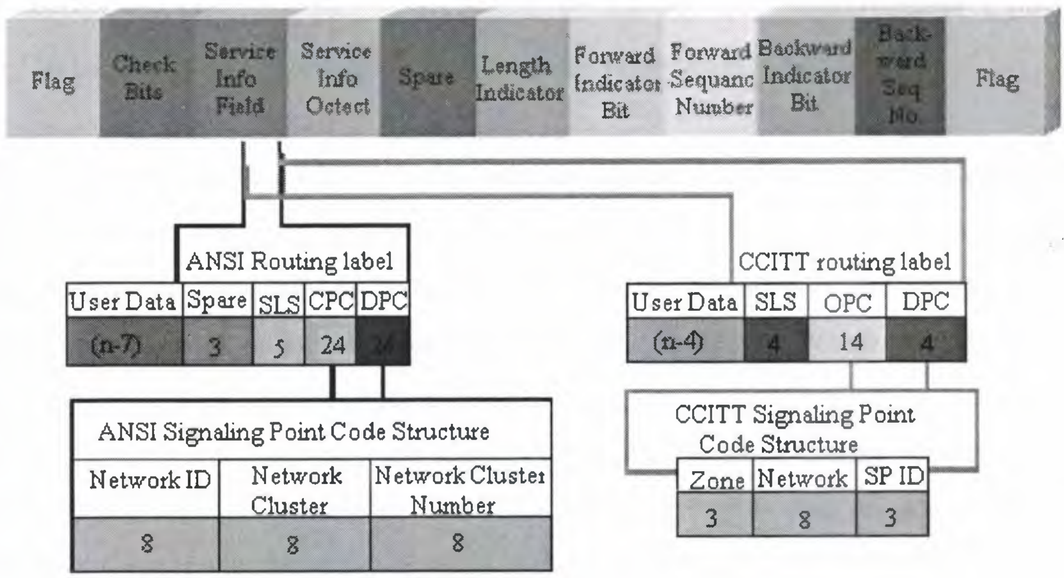

Figure 3.12 Signaling Information Field

- **Routing Label**

In the figure above there are two routing labels depicted. It should be made clear that both of these routing labels will not be found in the same MSU. The attempt made in the drawing is to illustrate an ANSI routing label and a CCITT (ITU-TS) routing label. The one to be found in an MSU will depend, of course, on which standard is being employed by the node sending the message. In North America, this will generally be ANSI. For ANSI networks the routing label is identified as being the first 56 bits in the SIF (Signaling Information Field). ANSI provides for a network location to be identified by a 24 bit code. The first eight bits of this code is given the name of Network ID. Signaling Point Codes, in general, follow the general schema introduced by the telephone numbering scheme known as the North American Numbering Plan. In that plan, each succeeding value generally represents a smaller geographical area (area code, exchange) until the last number represents an individual phone line.

For the SS7, the distinction is hierarchical rather than geographical. That is the network identifier identifies a network to which a point code belongs. This can be a code indicating a broad general network, or it can be a reserved code used to indicate that the identity is assigned to a group of commonly administered nodes which, together, do not qualify for full network status. With one of these reserved codes used as the network identifier, an indication is given that the next value (Network Cluster) will be used to identify the network to which the node belongs [9].

Thus the code can be used to identify smaller networks within a larger network and, finally, narrow down the addressing (Network Cluster Member) to address a single signaling point. The only reserved Network Cluster Member is the "empty" byte (eight zeros) which is used to identify an STP. The standards reserve this value for STP usage, but they don't compel its use.
What that means is that when you see a node with a Network Cluster Member value of zero, you know that the node is an STP. However, when you see a node with a Network Cluster Member code of other than zero, there is no guarantee that the node is not an STP.

The remaining field is the Signaling Link Selection Code. You may recall that different SCCP services require different handling of this code. SCCP services break down into four categories.

These services are connectionless or connection oriented, each of which is further categorized as requiring or not requiring in-sequence delivery. The MTP level 3 normally rotates the SLS and each new code will direct the message to a new available link. When the SCCP requires it, the
MTP stops rotating the code and each new message is directed to the same link, thereby guaranteeing in-sequence delivery.

For the ANSI routing label, the only thing remaining to be identified is the field labeled "User Data." This is the actual message. The letter "n" represents the total amount of data in the SIF. Since the routing label is always 56 bits (7 octets), "n-7" represents the size of the message (total octets in the field, less the routing label).
The differences in the CCITT (ITU-TS) routing label are relatively minor. The names of the signaling point code fields are changed (Zone instead of Network ID, Network instead of Network Cluster, and Signaling Point ID instead of Network Cluster Member). The ITU-TS standard also elected to allot a different number of bits to two of the three portions of the code (3-8-3 instead of 8-8-8). Fewer bits were also allotted to the SLS (4 instead of 5, and the ANSI 1996 standard supports 8 bits for the SLS).

And, finally, the computation of the message size based on the total data in the SIF changes because only four octets need to be subtracted from the total data.

One remaining field in the MSU is the Check Bits field. This is an algorithm that allows the MTP to determine whether the number of bits transmitted is the same as the number of bits received.

The last field is the trail flag. The trail flag is not used in the U.S. Therefore, in ANSI networks the MSU ends with the Check Bits.

3.4.2 The Link Status Signal Unit (LSSU)

As the name implies, the Link Status Signal Unit is used by the MTP to provide status information to the signaling points at either end of the link. These indications are placed in the one or two octet field called the Link Status field. Status indications relate to the current mode under which the MTP is monitoring the link. Recall that our earlier discussions about functionality mentioned two error rate monitors. The first was the AERM (Alignment Error Rate Monitor) employed during link alignment.

At the beginning of the alignment process (while in the **O**ut-of-alignment state) the MTP provides a status indication of "O." When the alignment has proceeded to the proving period, the status of "N" (for **N**ormal proving period) or the status of "E" (for **E**mergency proving period) is sent. By the way, as long as we are on the subject of Normal vs. Emergency alignment you may have wondered how the selection is made. The node itself does the selection, either as required by the application or by pre-configuration. The format of the Link Status Signal Unit is illustrated in figure 3.13.

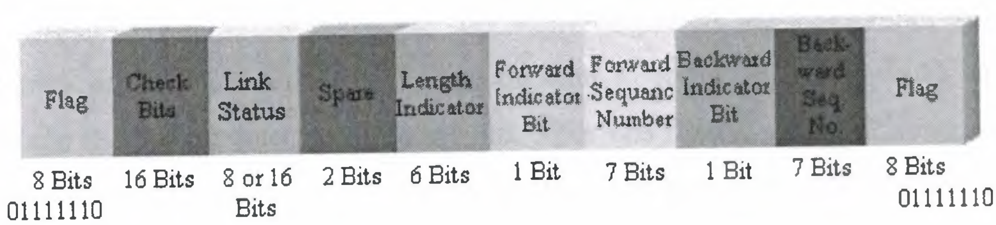

Figure 3.13 Link Status Signal Unit (LSSU)

When a link is first powered up, and before the "O" status indication, the "OS" (Out-of-Service) status is sent. The same status is sent any time the link can neither receive nor send MSUs for any reason other than a Processor Outage.

The sending of these status indications is somewhat of an interactive process on the part of the MTPs at either end of the link. For example, if the MTP at one end of the link has an "N" proving status it will send "N". If the other end sends an "E" status, the side with the Normal status will not change that status. However, it will not cause a delay in the link realignment by performing the longer proving period. It will, instead, perform the shorter "E" proving period to be consistent with the realignment being performed by the other side. Even as it does so, it still sends "N".

While the MTP is monitoring normal transmission (using the SUERM or Signal Unit Error Rate Monitor) it will send a status of "B" (link Busy). That same MTP stops sending either positive or negative acknowledgments. The MTP at the other end of the link sets a long timer to await the clearance of the congestion. It also resets the "excessive delay of acknowledgment" timer every time it receives (periodically during the congestion period) a new indication of "B."

A status of "PO" is sent when the MTP detects a problem in delivering messages to levels 3 or 4. The same MTP then begins message discard. When the opposite MTP receives "PO", it informs level 3 and places Fill In Signal Units on the link.

3.4.3 The Fill In Signal Unit (FISU)

If we pull a link out of the link port, the MTP will detect a "failed link." This is due to the fact that the MTP continues to monitor Signal Units and now has no data with which to deal. While nodes are communicating, there are many times when no messages are being sent. If there were no data to read, the MTP would have to conclude that it had detected a "failed link". For this reason, the MTP cannot allow any time period in which the link carries no data. The transmitting MTP fills all such blanks with Fill in Signal Units. The format of the Fill in Signal Unit is shown below in figure 3.14.

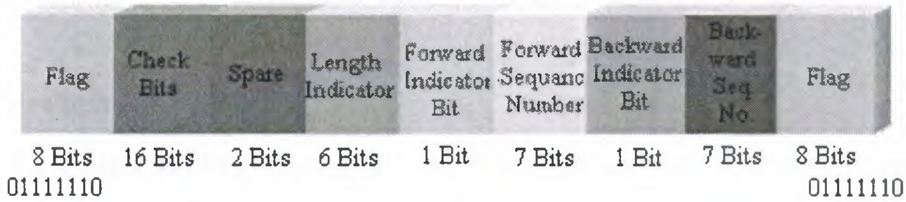

Figure 3.14 the Fill In Signal Unit (FISU)

As we can see, the format just became even simpler. The MTP can continue to monitor for valid flags, octet integrity and package size. It simply finds no data to send to level 3. On the transmitting side, no changes are made. The FSN applied to the last message simply reappears in each FISU until a new message is sent.

There is another time when the FISU becomes handy. During the alignment procedure the MTP uses the AERM to monitor the link. Until the link is placed in an In Service state, there are no messages being sent. Therefore, the MTP puts FISUs on the link and monitors them during the proving period.

3.5 Summary

In this chapter, we reviewed the background and the basic concepts of signaling system number 7. We discussed the SS7 network architecture and its essential components by describing its nodes and links. We introduced the protocol stack of this system by explaining each layer of information and its main function in the network. And finally we identified the signaling units of SS7 network to analyze the ways in which the protocol manages to deliver information from node to node and to the final destination of the message.

A comprehensive discussion of SS7 applications will be presented in Chapter four.

4. SS7 APPLICATIONS

4.1 Overview

The traditional SS7-based telephony networks are widely deployed around the entire globe. Whenever a new network is setup, it must provide connectivity with the existing networks. There are a wide variety of potential applications for SS7 technology. This chapter will provide a sampling of applications for which SS7 is used.

4.2 SS7 and Internet Protocol (IP) Signaling Systems

Public telephone systems today use the Internet for both; telephony voice and to carry signaling system 7 (SS7) messages. The Internet can provide reliable communication services by using the packet based transmission technologies used by IP-based protocols.

SS7 messages can be directly transported over IP networks or the functional equivalent of SS7 control message can be sent as control messages (e.g. text based messages) directly between elements connected to a data network (e.g. the Internet).

Internet telephone systems are primarily composed of media gateways (MGs) and one or more media gateway controllers (MGCs). When Internet telephone systems interconnect to other networks such as the Public Switched Telephone Network (PSTN), they use signaling gateways (SG) or Network Gateways (NGWs). Some of the more common IP Telephony Systems include session initiated protocol (SIP), media gateway control protocol (MGCP), MEGACO, and ITU's H.323 protocol.

Figure 4.1 shows that SS7 signaling systems can be interconnected with voice over data networks and that SS7 messages can be transported over the Internet protocol.

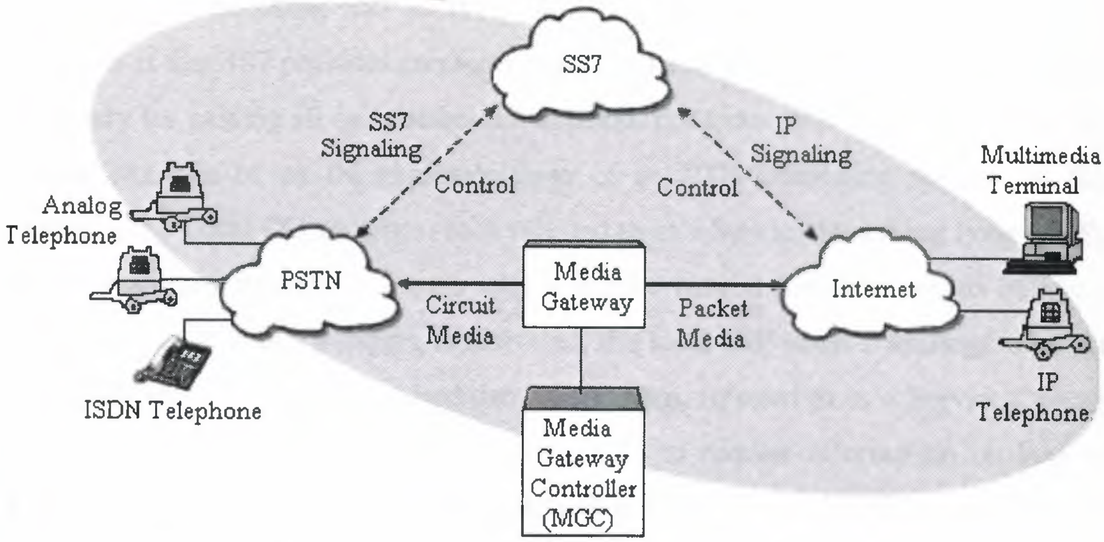

Figure 4.1 SS7 Interconnection with Voice over Data Networks

Figure 4.1 shows that analog and digital telephones are connected to the PSTN. To interconnect these telephones to voice over data network telephones, the media portion of each communication session is routed through a media gateway where it is converted from the PSTN circuit switched form to an IP packet data media format (packetized voice.) This diagram shows that the packet media can be routed through a data network (e.g. Internet) to an endpoint communication terminal such as a multimedia computer or an IP telephone. This diagram also shows that the SS7 network can control the PSTN through SS7 signaling messages and it can communicate to the media gateway through IP signaling messages.

4.3 SS7 and Intelligent Networking Applications

The Intelligent Network (IN) refers to architecture for implementing intelligence and advanced functionality within the telephone network. Also referred to as Advanced Intelligent Network (AIN) in the U.S., the IN architecture uses SS7 as the underlying data communications protocol [8].

The need for IN architecture has become increasingly important in recent years as customers have demanded more sophisticated services and as regulators. IN enables PSTN operators to implement a more uniform set of services (e.g., Caller ID) across a diverse installed base of central office switches and to provide new enhanced services such as 800# rerouting and one number follow me.

A key reason for using SS7 as the underlying network for implementing IN/AIN networks is that SS7 provides communications networks with the reliability and speed necessary for passing all call control information. For example, consider the relatively simple example of an IN implementation of an 800# translation service. In this example, the Local CO switches, each referred to as a Service Switching Point (SSP), are programmed to detect all calls which require special handling and to initiate a trigger function. When a trigger is activated, the local SSP sends a message over the SS7 network to a remotely located database system, referred to as a Service Control Point (SCP). This message is used by the SCP to request information on how to handle the call. Based on the trigger presented and the call characteristics, an SLP (Service Logic Program) running on the SCP determines the appropriate action and sends this information back to the SSP. The SSP then handles the call in the determined manner.

Even from this brief description, it is clear that the entire process must be very fast and highly reliable. If the process is too slow, there will be a noticeable delay; if the process is unreliable, the call will either be lost or handled in some incorrect manner. This example also demonstrates how IN, in using a centralized database(s) (the SCP), eliminates the need to update tables or programs in each of the local CO switches whenever a service or number changes.

Key elements of the Intelligent Network include (Figure 4.2):

- **Service Switching Points (SSP):** These signaling points are the originators and terminators of signaling messages such as local Central Offices (CO) or exchanges.

- **Service Control Points (SCP)**: These signaling points are typically databases, such as the Line Information Database (LIDB) in a wireline network or the HLR (Home Location Register)/VLR (Visitor Location Register) in a wireless network. The program running on the SCP which determines how the call should be handled is referred to as the Service Logic Program (SLP).

- **Signal Transfer Points (STP):** These signaling points are the SS7 packet switches, or routers, which route traffic through the SS7 network.

- **Intelligent Peripheral (IP):** These IN network elements provide services which facilitate customer interaction such as voice prompting, voice storage, and fax storage.

- **Adjunct:** These IN network elements provide customer service functions through the CO Switch.

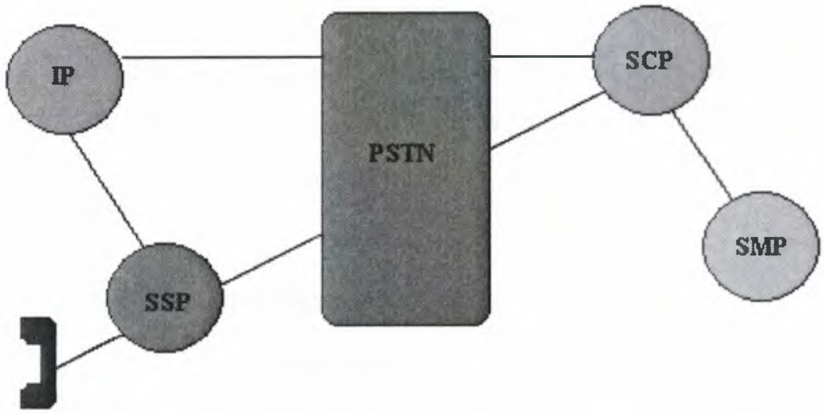

Figure 4.2 Classic Physical IN Architecture

4.4 Wireless Network Applications

Supporting the mobility of wireless subscribers is a key challenge in designing wireless systems. The wireless systems' signaling requirements are significantly more demanding than those needed to perform similar functions for land line systems. As a result, wireless networks must incorporate more advanced and comprehensive signaling and control systems. Even when the mobile unit is within its home network, tracking a mobile's location is required for terminating calls, authentication, and hand-off functions. The trend is for system operators to offer customers increasingly sophisticated levels of mobility, which in turn requires sophisticated signaling between systems to support roaming, registration and routing functions. SS7 is often the signaling system of choice.

SS7 is also widely used for signaling between wireless networks and the PSTN, and increasingly between the wireless network subsystems. In GSM networks, SS7 is also used for signaling between Base Station Controllers (BSC) and the Mobile Switching Center (MSC).

As shown in Figure 4.3, each system maintains its own HLR (Home Location Register) and VLR (Visitor Location Register) database. A subscriber's full record is maintained at a single HLR by the subscriber's home system. Records for subscribers who are only visiting the system are maintained in the VLR. When a subscriber travels to another service area and wishes to make a call, the visited system recognizes upon registration (when the mobile unit is first turned on) that the user is from another area and has a dialog with the subscriber's home system, using SS7 to temporarily register the visiting mobile unit in its VLR. In addition, the home system marks its HLR record for the subscriber so that calls to the subscriber are automatically routed to the visited location.

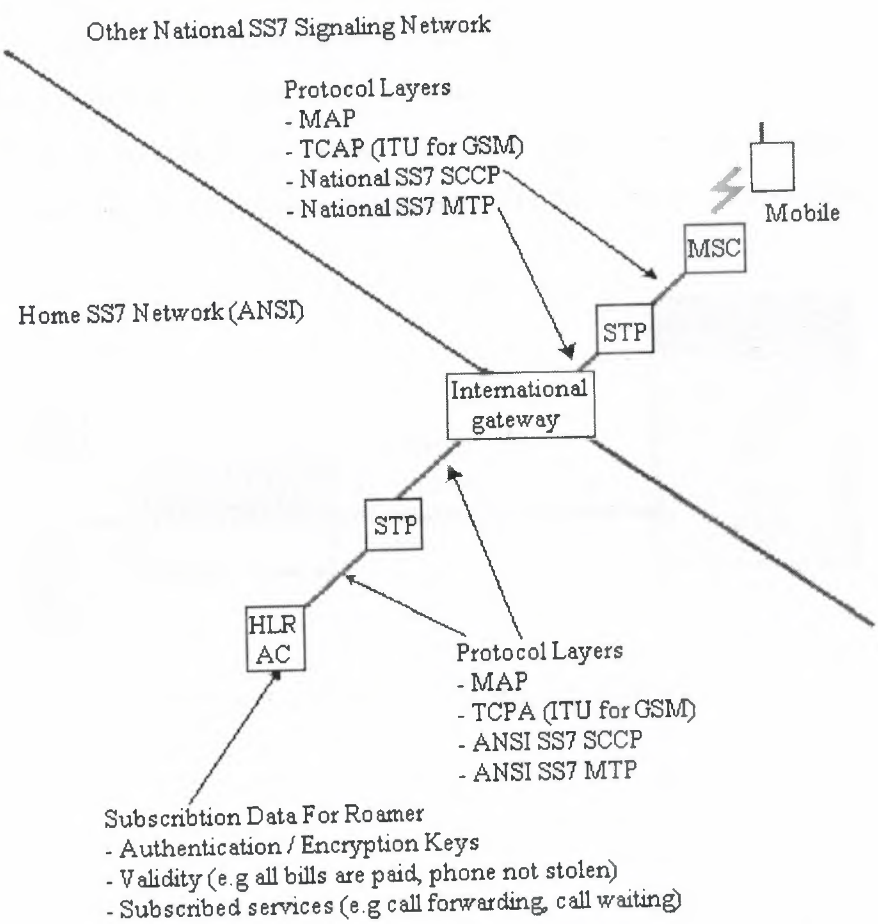

Figure 4.3 SS7 Supports Roaming in a Wireless Network

4.5 Interactive Voice Response (IVR) Applications

SS7 can be used with IVR systems to improve efficiency, provide new functionality, and reduce telecommunications costs. Specifically, the key benefits of using SS7 in an IVR application include [8]:

- Faster call set-up and call tear down.
- Increased information associated with a call (ANI/DNIS) intelligent routing.
- Ability to implement large distributed IVR systems.
- Ability to reduce telecommunications costs (e.g., 800# service) by purchasing high capacity trunks.
- Ease of upgrade to new services.

Figure 4.4 illustrates an example of a large distributed IVR application which uses SS7 signaling to route the call to the appropriate application. Note that an SS7 Gateway is used to concentrate traffic from two of the locations to reduce the number of SS7 links required. In determining whether to upgrade an IVR system with an SS7 capability, it is important to weigh the cost savings gained through increased efficiency and lower per call costs versus the additional cost of the SS7 links.

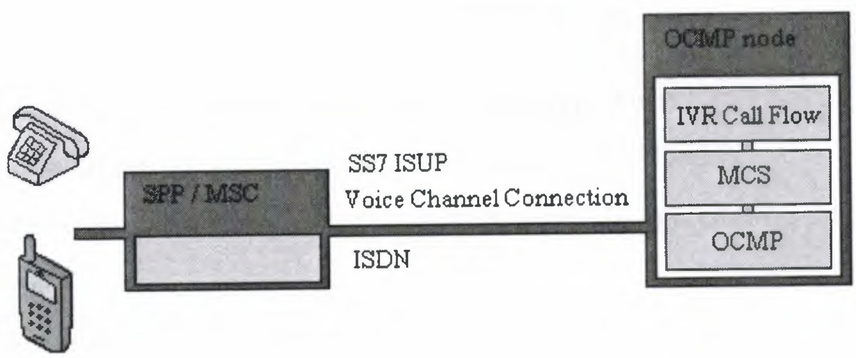

Figure 4.4 Network IVR Application Service

One of the challenges in implementing an IVR system is to ensure that either the call is routed to the location where the associated data resides, or that the data is passed to where the call is handled. This challenge is particularly important in large and/or distributed IVR systems which handle multiple applications. It is in these types of systems where SS7 is likely to be most applicable.

4.6 Call Center Applications

SS7 can also be used with call centers to improve efficiency, provide new functionality, and to reduce telecommunications costs. An outbound call center will typically use a predictive dialer, only routing answered calls to an agent. Several outgoing lines for each agent are provided to optimize agent utilization.

A key challenge in developing and operating these systems, however, is determining when a call is answered. Traditional telephone systems do not provide any specific signaling to indicate answer or hang-up. SS7 does, however, provide answer and hang-up indication to the call center, resulting in more efficient operation and lower operating costs.

Thus, key potential benefits of using SS7 in an outbound call center application include:

- Faster call set-up, answer detection and call tear down.
- Increased information associated with the call (ANI/DNIS).
- Ability to reduce telecommunications costs by purchasing high capacity trunks.
- Ease of upgrade to new services.

SS7 has many applications in call centers, particularly large call centers. For example, SS7 can be used to improve the efficiency of an inbound call center with intelligent routing and with faster ANI (Automatic Number Identification), or to implement a distributed call center by using SS7 to link in remote agent stations. SS7 can be used to reduce telecommunications costs for a large inbound call center/IVR application where calls are routed to the call center from a distributed IVR system. In this example, the remotely located IVR would transfer a call to the call center only when a live agent is available to take it, rather than immediately transferring all calls to the call center and incurring additional long distance telephone charges while the calls are queued.

As with the IVR application, the system designer must consider the cost savings gained through increased efficiency and lower per call costs versus the additional cost of the SS7 links when determining whether to upgrade a call center with SS7 capability [10].

4.7 ATM – SS7 Inetworking

Major carriers are rapidly deploying ATM (Asynchronous Transfer Mode) in both public and private networks as a result of simplicity, flexibility, and ability to support high-bandwidth applications efficiently. Even the third generation networks, which present a uniform view to both mobile and fixed subscribers, have also been widely using ATM as an underlying technology.

ATM has emerged as one of the key technologies in such a multi-service network. ATM is a cell-based (small, fixed-size structure) technology employing statistical multiplexing techniques to allow full usage of transport bandwidth. From its inception, ATM has been designed to optimize use of high-bandwidth optical cables. It can efficiently carry almost any type of traffic requiring any amount of bandwidth, thus enabling the introduction of a wide spectrum of services at a competitive price. ATM networks represent a milestone in the convergence of voice and data networks.

4.7.1 Need for ATM- SS7 Interworking

In spite of growth in ATM networks, SS7 networks will continue to exist. Among the major reasons for this are huge investments hat have been made in these networks and the fact that in a large number of geographical areas around the world little need exist for services other than the basic ones.

It also possible that ATM will be introduced only in the toll network to maintain the stability of the services offered by SS7- based networks. It is evident that even as ATM networks come up, SS7-based narrowband networks will continue to exist and grow.

This can result in primarily two types of network configurations in which interworking between these two networks are required. In one of these scenarios, SS7- based narrowband networks are connected through an ATM backbone network. This backbone ATM network could possibly be owned by an independent carrier, as shown in figure 4.5.

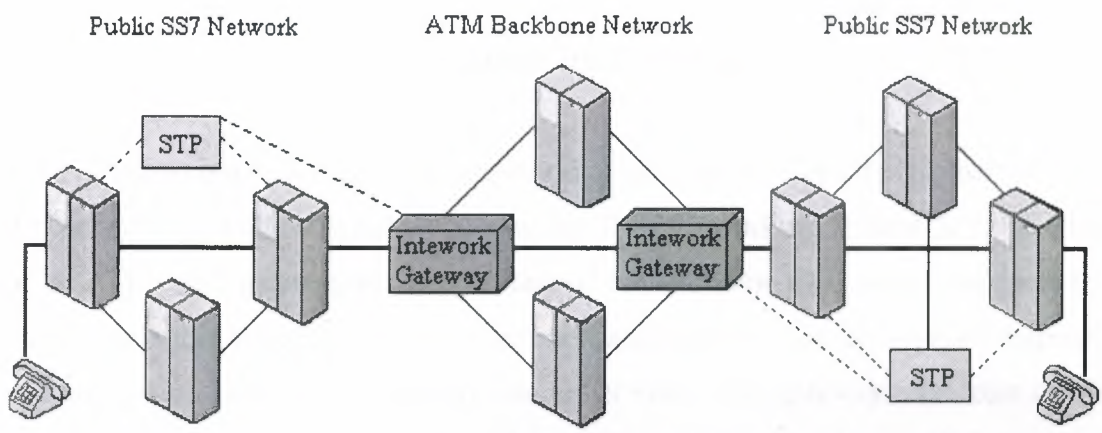

Figure 4.5 Interworking Scenario A

In the second scenario, an ATM network is set up by a service provider in a local area while the long- distance carrier continues to be SS7- based narrowband as shown in figure 4.6. The ATM network in question could be a private or a public network providing ATM services in a local area.

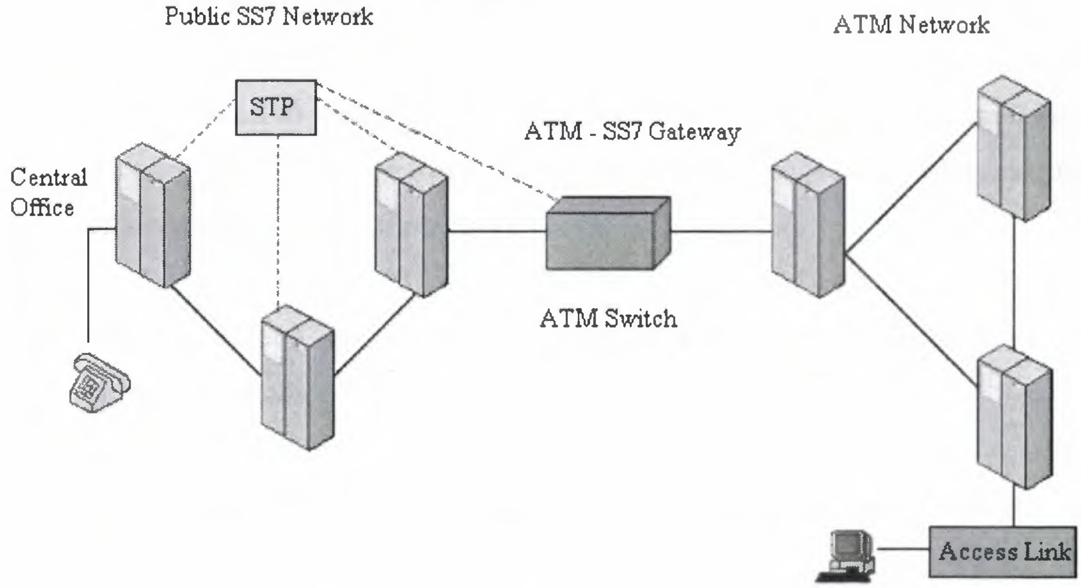

Figure 4.6 Interworking Scenario B

Both of these scenarios require an interworking solution to bridge the gap between the two networks and allow seamless integration. The interworking solution is in the form of an ATM- SS7 gateway at the boundary of the two networks, which interprets the traffic and signaling going in either direction and performs the appropriate mapping and conversion between the networks on either side. The gateway takes care of the differences between two networks such as differences in signaling, addressing, and transport technology. The gateway can either be located at the edge of the ATM network (in an ATM switch) or as a stand- alone system between the edges of the two networks.

Some of the important requirements for such an interworking solution are the following:

- Bearer service interworking should be possible between the two networks. As the services provided in the narrowband SS7 network are a subset of the services provided in the ATM, the services offered across the two networks will be limited to the narrowband circuit-mode services.

- The interworking should take place transparently. A user in the ATM network is not expected to invoke any special procedure or install additional hardware to place a call to a narrowband subscriber. Similarly, a user in the narrowband network should be able to place calls to the ATM subscriber using existing procedures and user equipment.

- There should be no loss in service quality when the call transist from one network to another.

- The existing SS7 and ATM networks should not require any hardware and software upgrades or changes.

4.7.2 ATM- SS7 Gateway

In the configuration shown in Figure 4.7, both of the networks attached to the ATM- SS7 gateway assign it an address (for routing based on the network format used). The gateway will have a SS7 point code and an ATM address to identify the points of attachment to the narrowband SS7 network and ATM network, respectively. The routing tables in the two networks are configured in such a way that call request (signaling) originating in the narrowband and destined for the ATM network or vice versa gets routed to the gateway.

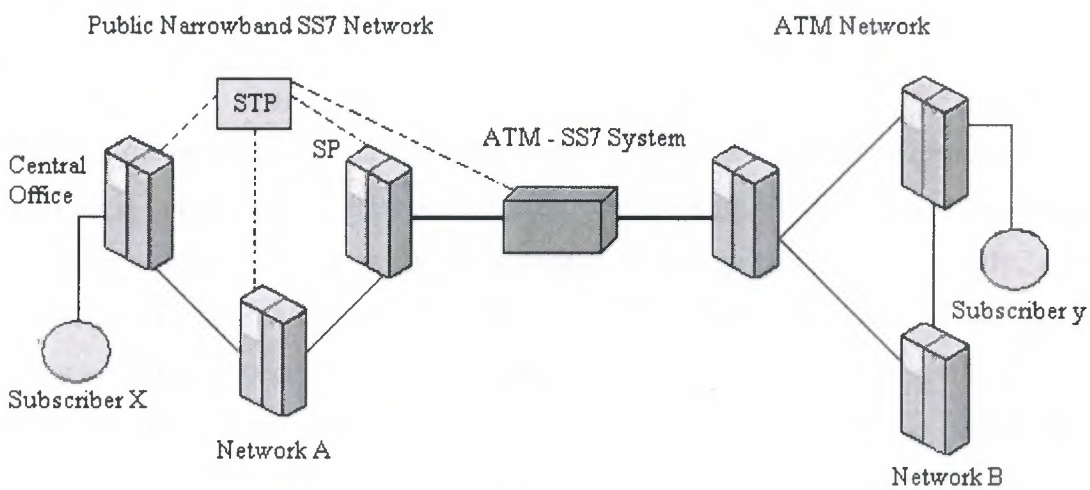

Figure 4.7 Reference Network Scenario for ATM- SS7 Interworking

The basic operation of the gateway can be best explained by describing the scenario of a connection set up across the two subscribers shown in Figure 4.7: subscriber X in the SS7- based narrowband network A and subscriber Y, the call request (signaling) is routed through a series of signaling transfer points (STPs) and signaling points (SPs) in the narrowband SS7 network to the gateway, allocating circuit resources along the path.

When call request reaches the gateway, a routing table lookup at the gateway determines the network where subscriber Y lies (network B in figure 4.7). It also determines the signaling protocol and other parameters to be used on network B, including addressing method. It performs a mapping of the received call setup request (an SS7 signaling message) from network B. To preserve service integrity, the gateway keeps as much signaling as information intact as is possible across the two networks.

Bearer service mapping is then performed at the gateway to map the bearer service requested by subscriber X in SS7 (network A), such as voice and 3.1-kHz audio, to that in ATM (network B). Based on the requested bearer service, appropriate QoS and traffic parameters are chosen. Also, as the called subscriber Y in the ATM network may be addressed in a way different from subscriber X, address translation is done to map the addressing schemes. This request is then sent across to the ATM switch in network B. the ATM switch then routes it further to the destination switch subscriber Y.

The user traffic starts flowing in either direction when the signaling sequence to set up the call is completed. At this point, traffic interworking between incoming and the outgoing streams of user traffic is activated at the gateway. Subsequently, when subscriber Y (or subscriber X) initiates release of the call, it is indicated to the gateway by the respective signaling protocol. The gateway then initiates release of the call in the other network. Interworking is a continuous function and is performed at every point of the call duration, from connection setup request to user traffic flow and subsequent call release.

A call setup request in the broadband to narrowband direction is treated in a similar manner at the gateway. However, only those calls from the ATM network, in which a narrowband service is requested, are allowed to go through by the gateway. Including the narrowband information elements in the signaling messages for call setup indicates such calls. This is not a restriction but a guideline facilitating seamless interworking between narrowband and ATM subscribers. If a broadband service is requested, the call setup is rejected be the gateway with an appropriate cause.

4.7.3 Functions of the ATM- SS7 Gateway

The main functions of an ATM-SS7 gateway can be broadly categorized as the following:

- Control and signaling interworking
- Traffic interworking

4.7.3.1 Control and Signaling Interworking

This refers to the capability of the gateway to map control and signaling between the two networks. It involves signaling interworking, service interworking, routing, and addressing. Each of these is now discussed in detail.

1. **Signaling Interworking**

Narrowband SS7 networks use narrowband integrated services digital network (ISDN) user part (N-ISUP) as the protocol for signaling between the SS7 nodes. Public ATM networks widely use broadband ISDN user part (B-ISUP) protocol for signaling while their private counterparts use private network-to-node interface (P-NNI) signaling for settings up and tearing down calls. On the ATM user interface, user-to-network interface (UNI) protocol is most widely used. These protocols are essentially message-based, with their contents defined by the standard organizations.

The information in the protocol messages is interpreted by the nodes and acted on according to defined procedures. This may involve sending out a new message to another node, feeding a tone, or performing any other action required be the procedures.

The signaling messages and procedures used to set up and tear down calls are different in ATM and SS7 networks. More apparent is the difference in the information carried in signaling messages. Thus, one of the important functions of the gateway is to perform signaling interworking (i.e., to map the signaling messages, information elements, and procedures used in one network to the corresponding messages, information elements, and procedures used in the other network).

2. **Bearer Service Interworking**

Bearer service interworking at the gateway allows bearer services supported by the ATM network to be mapped transparently to those in SS7 network and vice versa. Bearer services supported by the SS7 network are limited to circuit-mode services requiring bearer capabilities such as voice, 3.1-kHz audio and n x 64 unrestricted. Bearer capability for services in ATM networks is defined in terms of cell rate and type of traffic (such as constant bit rate [CBR] and variable bit rate [VRB]). The gateway is responsible for mapping bearer capabilities between SS7 and ATM. Table 4.1 shows the Bearer Services Interworking.

Table 4.1 Bearer Services Interworking

Bearer Service on SS7	**Bearer Service on ATM**
Speech	• CBR- based native mode ATM connection (without any AAL), effectively emulating a 64-kbps connection • CBR connection using AAL_1, effectively emulating a 64-kbps connection • VBR- based AAL_2 connection with silence detection, suppression, and voice encoding schemes such as 16 kbps and 12.8 kbps based on International Telecommunications Union- Telecommunications (ITU-T) Recommendation G.728 • CBR connection using ALL_5, effectively emulating a 64-kbps connection
3.1-kHz audio	• CBR- based native mode ATM connection (without any AAL), effectively emulating a 64-kbps connection • CBR connection using AAL_1, effectively emulating a 64-kbps connection • VBR-based AAL_2, connection with a standard modem

	tone detection and encoding scheme (if it is know in advance through signaling procedures that the call is a modem/fax call) • CBR connection using AAL_5, effectively emulating a 64-kbps connection
n x 64 unrestricted	• CBR connection using AAL_1, with end-to-end timing • CBR connection using AAL_2, with end-to-end timing

As is clear from Table 4.1, there are several possibilities on the ATM network interface for mapping a given bearer capability on the SS7 interface. For a given narrowband service, the selection of ATM QoS parameters and the ATM adaptation layer (such as ATM adaptation layer 1 [AAL_1], AAL_2, or AAL_5) can depend on several factors, including the following:

- Capabilities of the destination/incoming ATM network
- Capabilities of the ATM hardware in gateway
- Timing arrangements between the two networks

The advantage of using AAL_2 as the adaptation layer is that it allows for voice compression and silence detection and removal. AAL_5 can also be used, as it is required for signaling and thus present in virtually every ATM node. AAL_1 offers an advantage in terms of simplicity and interoperability between vendors.

In the reverse direction, only circuit-mode services from an ATM network can be interworked with those in SS7 network. If a broadband-specific service is requested, interworking is not possible and the call is rejected.

3. Address Mapping

Each end point in a network should be assigned an address to allow calls to and from it. In an ATM network, the address can be either a 20-byte ATM end system address (AESA) or a native E.164 address. Public ATM networks typically use E.164 format, as it is supported be the existing public network infrastructure. Private ATM networks may use either format. On the other hand, SS7 networks typically route based on E.164 address. If an ATM subscriber requests a narrowband service, the end-user may be an ATM subscriber or a narrowband user. Accordingly, the address of the called party may be specified as an AESA or an E.164 address.

When this call is routed to the gateway, it should be converted to an E.164 address that can be routed by the SS7 network. This conversation can be done by using a local back end in the gateway (for small networks) or by using the services of an address resolution server (for large networks). In the reverse direction, an E.164 address received from the SS7 network may need to be mapped to an AESA address using any of the schemes mentioned earlier.

4. **Routing**

The SS7-ATM gateway should support call-by-call routing, capability to support switched operation. The gateway should have an interface to the routing in order to request an outgoing path for the incoming call. The request to the routing module includes information such as called party number, resource type requested, and implementation- specific information. Routing results in the selection of an outgoing facility (e.g., a narrowband trunk group or ATM virtual connection) on which to route the call.

5. **Connection Admission Control (CAC)**

The gateway should perform local connection admission control (CAC) function. The local CAC function should be able to determine, based on the traffic and QoS parameters requested in the call setup, whether setting up the connection violates the QoS guarantee of the established connections. It also finds out if enough resources are available to accommodate the call. If it reports failure, the gateway should reject the connection setup request.

6. **Operations, Administration, and Maintenance Functions**

The operations, administration, and maintenance (OAM) functions, which should be performed by the gateway toward the ATM networks, include defect and failure detection, fault localization, and system information. On the SS7 network interface, alarms related to physical transmission failures and interruptions should be handled by the gateway. Appropriate treatment should be given to the calls by the gateway in a failure or loss of continuity is detected on any of the interfaces. The gateway, being a critical pint in the network, should have high-availability. This involves the capability of the gateway to handle both faults and failures.

7. **Billing and Charging**

In circuit-switched networks, the charging plan is well defined and based on the time for which the connection is active, given that the entire bandwidth is reserved for the user. However, in ATM networks, bandwidth is shared; thus, a similar charging plan would be inappropriate. Charging in ATM networks has not yet fully evolved. It could be based on several factors, such as the type of QoS guarantee, port-based charging, or number of cells transferred successfully. As these networks become integrated, charging becomes more complex, and bilateral agreements should be established between the owners of the two networks.

4.7.3.2 Traffic Interworking

1. **TDM-ATM Conversation**

The transfer technology in ATM is cell-based (fixed-size packets of 53 bytes). An ATM connection can be viewed as a stream of cells arriving at a mean interval, depending on the type of traffic. At the receiving end (AAL), the cells are adapted to the type of application.

Transfer technology in time division multiplex (TDM) is based on circuit switching, which can be viewed as a stream of octets arriving at a fixed interval. The octets arrive at a fixed time rate of 8,000 times per second, achieving a bit rate of 64kpbs.

The TDM-ATM Bridge performs the conversion of TDM-encoded octets to ATM cells in the physical plane under the control of bearer service interworking, depending on the specified type of AAL, cell rate, and QoS parameters. These adaptations take place in hardware under software control. They may also include voice activity detection (VAD) and voice compression capabilities to support efficient bandwidth usage using AAL_2. Furthermore, they may support fax modulation and demodulation standards, again to allow efficient usage of bandwidth.

2. Echo Control and End-to-End Delay Handling

Echo is normally caused be feedback from end-user devices (acoustical feedback) and hybrids in the network. Delay is introduced into an end-to-end connection by factors such as packetization, compression algorithms, physical transmission time, and queuing delays at the ATM switches in the path. Echo signals become objectionable with increased delays and feedback levels. In SS7 and ATM networks, echo is removed by inserting echo cancellers in the voice path, which electronically remove the echo from voice signals.

The gateway should be able to identify whether or not echo control is required. The information that influences this decision includes address information, nature of circuit, signaling information received about echo control, bearer capability requested, and propagation delay information. The gateway then uses echo control-logic procedures to analyze the available information related to echo-control requirements in order to optimize the locations at which echo-control devices are provided in the connection. Echo control logic is involved when the bearer capability information indicates speech or 3.1-Khz audio.

3. Synchronization of Timing Sources

Both SS7 and ATM networks may be timed from separate sources. These sources may or may not be highly accurate in nature. Even with highly accurate timing sources, minor drifts in frequency and phase may occur. This can result in overflow or underflow of buffers for voice and data packets. The getaway should be able to absorb these differences by providing jitter and wander buffers of an appropriate size.

Depending on the location of the gateway and bilateral agreement between the two operators of networks, the gateway may use a simpler scheme in which only one source of timing is used by both of the networks. Some services Such as nx64 structured, require end-to-end timing all trough the network and between the two user terminals. The gateway should take this into account while such calls are being routed.

4.8 Summary

In this chapter, we discussed the main applications of signaling system number 7. We represented SS7 applications with IP protocols, intelligent networking, wireless networks, call center and interactive voice response systems. And finally we introduced the SS7-ATM interworking by identifying ATM signaling and the main building blocks of this system.

5. SS7 & VoIP INTERWORKING

1.5 Overview

Although voice over IP (VoIP) has been in existence for some years, service demands are forcing a rapid evolution of the technology. The pace of service integration (convergence) with new and existing networks continues to increase as VoIP products and services develop. Also, the promise of broadband services and the integration of voice and data at all levels further the need for VoIP applications.

Though VoIP is still evolving, packet-based telephony is becoming more advanced. Voice protocols have further developed to offer a richer set of features, scalability, and standardization than what was available only a few years ago [11].

5.2 Why VoIP

Telephone companies offload voice calls from public switched telephone networks (PSTNs) to voice-over-Internet Protocol (VoIP) networks because it is cheaper to carry voice traffic over Internet Protocol (IP) networks than over switched circuit networks. In the future, IP telephony networks are expected to enable innovative new multimedia services while working seamlessly with legacy telephone networks.

A VoIP network carries voice traffic cheaper than a switched circuit telephone network because IP telephony networks make better use of available bandwidth. In a public switched telephone network, for example, a dedicated 64 kilobits per second (kbps) end-to-end circuit is allocated for each call. In a VoIP network, digitized voice data is highly compressed and carried in packets over IP networks. Using the same bandwidth, a VoIP network can carry many times the number of voice calls as a switched circuit network with better voice quality. The savings realized in using VoIP networks are often passed onto users in the form of lower costs.

In addition to voice data, signaling data is exchanged between switched circuit telephone networks and VoIP networks. Signaling information is used to setup, manage and release voice calls, and support telephony services such as caller ID, toll-free calling, and mobile authentication and roaming services.

5.3 VoIP Functions

VoIP components must be able perform the same features as the PSTN network.

5.3.1 Signaling

Signaling in a VoIP network is just as critical as it is in the legacy phone system. The signaling in a VoIP network activates and coordinates the various components to complete a call. Although the underling nature of the signaling is the same, there are some technical and architectural differences in a VoIP network [3].

Signaling in a VoIP network is accomplished by the exchange of IP datagram messages between the components. The format of these messages is covered by any number of standard protocols. Regardless of which protocol and product suites those are used, these message streams are critical to the function of a voice-enabled network and might need special treatment to guarantee their delivery.

5.3.2 Database Services

Database services are a way to locate an endpoint and translate the addressing that two (Usually heterogeneous) networks use. For example, the PSTN uses phone numbers to identify endpoints, while a VoIP network could use an IP address (address abstraction could be accomplished with DNS) and port numbers to identify an endpoint. A call control database contains these mappings and translations. Another important feature is the generation of transaction reports for billing purposes. We can employ additional logic to provide network security, such as to deny a specific endpoint from making overseas calls on the PSTN side. This functionality, coupled with call state control, coordinates the activities of the elements in a VoIP network.

5.3.3 Call Connect and Disconnect (Bearer Control)

The connection of a call is made by two endpoints opening communications sessions between each other. In the PSTN, the public (or private) switch connects logical DS-0 channels through the network to complete the calls. In a VoIP implementation, this connection is a multimedia stream (audio, video, or both) transported in real time. This connection is the bearer channel and represents the voice or video content being delivered. When communication is complete, the IP sessions are released and optionally network resources are freed.

5.3.4 CODEC Operations

Voice communication is analog, while data networking is digital. The process of converting analog waveforms to digital information is done with a coder-decoder (CODEC, which is also known as a voice coder-decoder [VOCODER]). There are many ways an analog voice signal can be transformed, all of which are governed by various standards. The process of conversion is complex and beyond the scope of this paper. Suffice to say that most of the conversions are base on pulse coded modulation (PCM) or variations. Each encoding scheme has its own history and merit, along with its particular bandwidth needs.

In addition to performing the analog to digital conversion, CODECs compress the data stream, and provide echo cancellation. Compression of the represented waveform can afford you bandwidth savings. The bandwidth savings for the voice services can come in several forms and work at different levels. For example, analog compression can be part of the encoding scheme (algorithm) and does not need further digital compression from the higher working layers of the media gateway application. Another way to save bandwidth is the use of silence suppression, which is the process of not sending voice packets between the gaps in human conversations.

Using compression and/or silence suppression can result in sizable bandwidth savings. However, there are some applications that could be adversely affected by compression. One example is the impact on modem users. Compression schemes can interfere with the functioning of modems by confusing the constellation encoding used. The result could be modems that never synchronize or modems that exhibit very poor throughput. Some gateways might implement some intelligence that can detect modem usage and disable compression. Another potential issue deals with low-bit-rate speech compression schemes, such as G.729 and G.723.1. These encoding schemes try to reproduce the subjective sound of the signal rather than the shape of the waveform. A greater amount of packet loss or severe jitter is more noticeable than that of a non compressed waveform. However, some standards might employ interleaving and other techniques that can minimize the effects of packet loss.

The output from the CODECs is a data stream that is put into IP packets and transported across the network to an endpoint. These endpoints must use the standards, as well as a common set of CODEC parameters. The result of using different standards or parameters on both ends is unintelligible communication. Table 5.1 lists some of the more important encoding standards covered by the International Telecommunications Union (ITU). As we can see, there is a price paid for reduced bandwidth consumption by increased conversion delay.

Table 5.1 ITU Encoding Standards

ITU Standard	Description	Bandwidth (Kbps)	Conversation Delay (ms)
G.711	PCM	64	< 1.00
G.721	ADPCM	32, 16, 24, 40	< 1.00
G.728	LD-CELP	16	~ 2.50
G.729	CS-ACELP	8	~ 15.00
G.723.1	Multirate CELP	6.3, 5.3	~ 30.00

5.4 Signaling in VoIP Networks

VoIP networks carry SS7-over-IP using protocols defined by Signaling Transport (sigtran) working group of the Internet Engineering Task Force (IETF), the international organization responsible for recommending Internet standards. The sigtran protocols support the stringent requirements for SS7/C7 signaling as defined by International Telecommunication Union (ITU) Telecommunication Standardization Sector [11]. In IP telephony networks, signaling information is exchanged between the following functional elements.

5.4.1 Media Gateways

Media gateways are responsible for call origination; call detection, analog-to-digital conversion of voice, and creation of voice packets (CODEC functions). In addition, media gateways have optional features, such as voice (analog and/or digital) compression, echo cancellation, silence suppression, and statistics gathering.

The media gateway forms the interface that the voice content uses so that it can be transported over the IP network. Media gateways are the sources of bearer traffic. Typically, each conversation (call) is a single IP session transported by a Real-time Transport Protocol (RTP) that runs over UDP.

Media gateways exist in several forms. For example, media gateways could be a dedicated telecommunication equipment chassis, or even generic PC running VoIP software. Their features and services can include some or all of the following.

- Trunking gateways that interface between the telephone network and a VoIP network. Such gateways typically manage a large number of digital circuits.

- Residential gateways that provide a traditional analog interface to a VoIP network. Examples of residential gateways include cable modem/cable set-top boxes, xDSL devices, and broadband wireless devices.

- Access media gateways that provide a traditional analog or digital PBX interface to a VoIP network. Examples include small-scale (enterprise) VoIP gateways.

- Business media gateways that provide a traditional digital PBX interface or an integrated soft PBX interface to a VoIP network.

- Network access servers that can attach a modem to a telephone circuit and provide data access to the Internet.

- Discreet IP telephones units.

5.4.2 Media Gateway Controllers

Media gateway controllers house the signaling and control services that coordinate the media gateway functions. Media gateway controllers could be considered similar to that of H.323 gatekeepers. The media gateway controller has the responsibility for some or all of the call signaling coordination, phone number translations, host lookup, resource management, and signaling gateway services to the PSTN (SS7 gateway). The amount of functionality is based on the particular VoIP enabling products used.

In a scalable VoIP network, we can breakup the role of a controller into signaling gateway controller and media gateway controller. For calls that originate and terminate within the domain of the VoIP network, only a media gateway controller might be needed to complete calls. However, a VoIP network is frequently connected to the public network. We could use a signaling gateway controller to directly connect to the SS7 network, while also interfacing to the VoIP network elements. This signaling controller would be dedicated to the message translation and signaling needed to bridge the PSTN to the VoIP network.

The services of these devices are defined by the protocols and software they are running. There are several protocols and implementations that any number of vendors could deploy. Knowing the details of how the devices use their suite of protocols is important to designing the IP backbone that is to service the VoIP elements.

5.4.3 IP Network

We can view the VoIP network as one logical switch. However, this logical switch is a distributed system, rather than that of a single switch entity; the IP backbone provides the connectivity among the distributed elements. Depending on the VoIP protocols used, this system as a whole is sometimes referred to as a (softswitch architecture). Figure 5.1 shows an example of a VoIP network configuration.

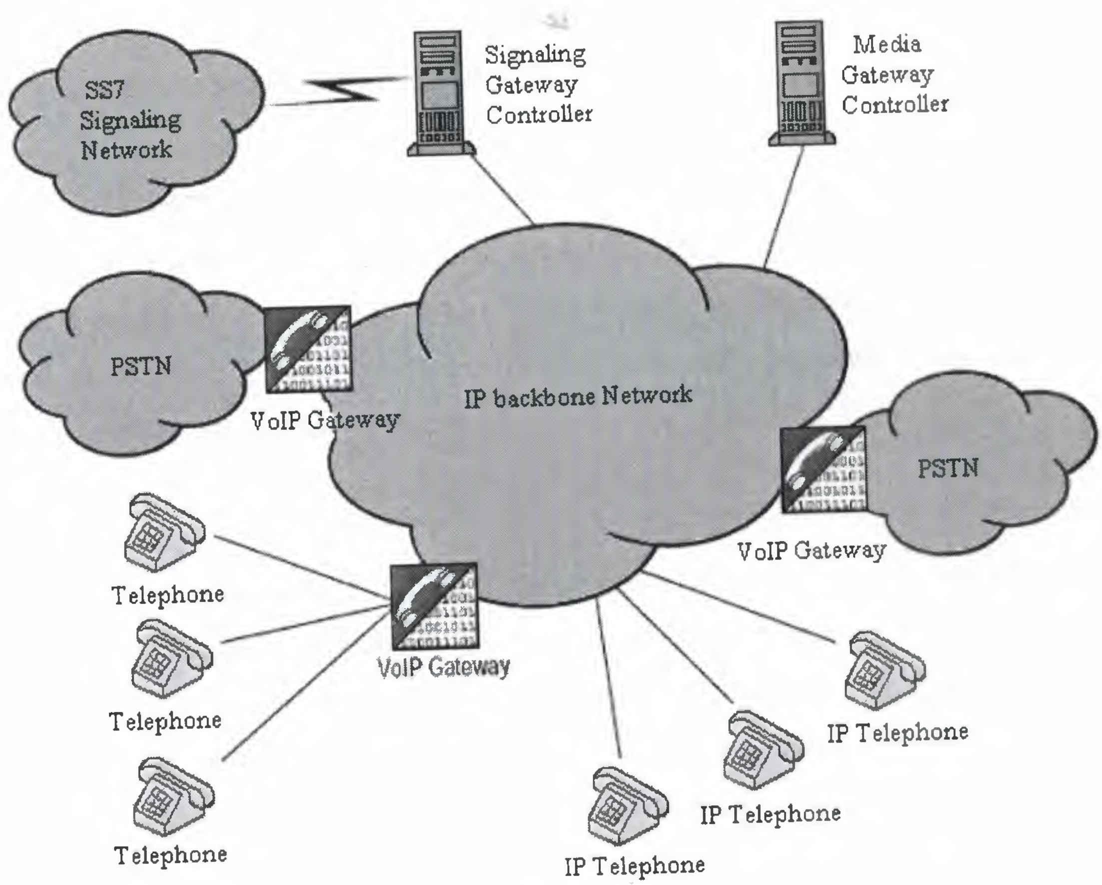

Figure 5.1 Full Service VoIP Network

The IP infrastructure must ensure smooth delivery of the voice and signaling packets to the VoIP elements. Due to their dissimilarities, the IP network must treat voice and data traffic differently. If an IP network is to carry both voice and data traffic, it must be able to prioritize the different traffic types.

There are several correlations to the VoIP and circuit-switching components, however there are many differences. One is in the transport of the resulting voice traffic. Circuit-switching telecommunications can be best classified as a TDM network that dedicates channels, reserving bandwidth as it is needed out of the trunk links interconnecting the switches. For example, a phone conversation reserves a single DS-0 channel, and that end-to-end connection is used only for the single conversation. IP networks are quite different from the circuit-switch infrastructure in that it is a packet-network, and it is based on the idea of statistical availability. Class of service (CoS) ensures that packets of a specific application are given priority. This prioritization is required for real-time VoIP applications to ensure that the voice service is unaffected by other traffic flows.

5.5 VoIP Service Considerations

VoIP traffic has a number of issues must be carefully considered, such as traffic parameters and network design. Without such due diligence, we could be faced with service that does not function reliably or is severely degraded. These important considerations are as follows.

5.5.1 Latency

Latency (or delay) is the time that it takes a packet to make its way through a network end to end. In telephony terms, latency is the measure of time it takes the talker's voice to reach the listener's ear. Large latency values do not necessarily degrade the sound quality of a phone call, but the result can be a lack of synchronization between the speakers such that there are hesitations in the speaker' interactions.

Generally, it is accepted that the end-to-end latency should be less than 150 ms for toll quality phone calls. To ensure that the latency budget remains below 150 ms, we need to take into account the following primary causes of latency. When designing a multi-service network, the total delay that a signal or packet exhibits is a summation of all the latency contributors.

5.5.2 Jitter

Jitter is the measure of time between when a packet is expected to arrive to when it actually arrives. In other words, with a constant packet transmission rate of every 20 ms, every packet would be expected to arrive at the destination exactly every 20 ms. this situation is not always the case. For example, Figure 5.2 shows packet one (P1) and packet three (P3) arriving when expected, but packet two (P2) arriving 12 ms later than expected and packet four (P4) arriving 5 ms late.

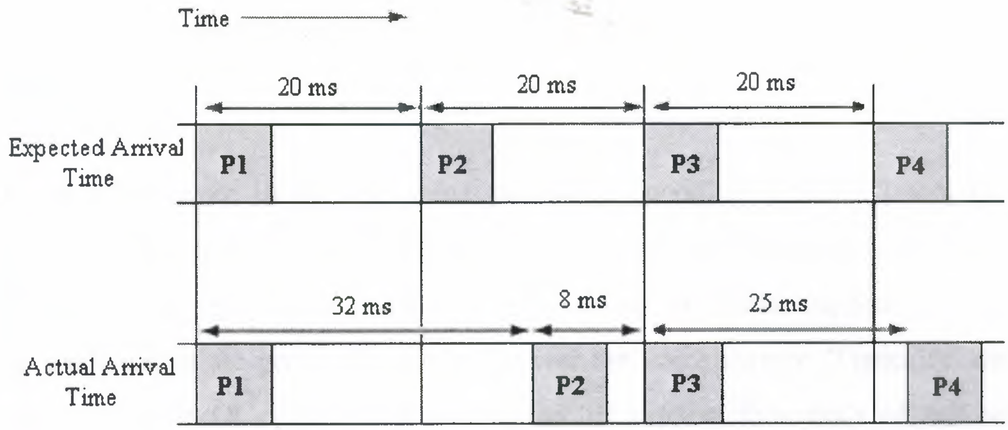

Figure 5.2 Example of Jitter

The greatest culprit of jitter is queuing variations caused by dynamic changes in network traffic loads. Another cause is packets that might sometimes take a different equal-cost link that is not physically (or electrically) the same length as the other links.

Media gateways have play-out buffers that buffer a packet stream so that the reconstructed voice waveform is not affected by packet jitter. Play-out buffers can minimize the effects of jitter, but cannot eliminate severe jitter.

Although some amount of jitter is to be expected, severe jitter can cause voice quality issues because the media gateway might discard packets arriving out of order. In this condition, the media gateway could starve its play-out buffer and cause gaps in the reconstructed waveform.

5.5.3 Bandwidth

We can determine how much bandwidth to set aside for voice traffic using simple math. However, in a converged voice and data network, we have to make decisions on how much bandwidth to give each service. These decisions are based on careful consideration of our priorities and the available bandwidth we can afford. If you allocate too little bandwidth for voice service, there might be unacceptable quality issues. Another consideration is that voice services are less tolerant to bandwidth depletion than that of Internet traffic. Therefore, bandwidth for voice services and associated signaling must take a priority over that of best-effort Internet traffic.

If a network were to use the same prevailing encoding (CODEC) scheme as the current PSTN system, bandwidth requirements for VoIP networks would tend to be larger than that of a circuit-switched voice network of similar capacity. The reason is the overhead in the protocols used to deliver the voice service. Typically, we would need speeds of OC-12c/STM-4 and higher to support thousands of call sessions. However, VoIP networks that employ compression and silence suppression could actually use less bandwidth than a similar circuit-switched network. The reason is because of the greater granularity in bandwidth usage that a packet-based network has in comparison to a fixed, channel size TDM network.

Allocations of network bandwidth are based on projected numbers of calls at peak hours. Any over-subscription of voice bandwidth can cause a reduction in voice quality. Also, we must set aside adequate bandwidth for signaling to ensure that calls are complete and to reduce service interruptions.

5.5.4 Packet Loss

Packet loss occurs for many reasons, and in some cases, is unavoidable. Often the amount of traffic a network is going to transport is underestimated. During network congestion, routers and switches can over flow their queue buffers and be forced to discard packets. Packet loss for non-real-time applications, such as Web browsers and file transfers, are undesirable, but not critical. The protocols used by non-real-time applications, usually TCP, are tolerant to some amount of packet loss because of their retransmission capabilities.

Real-time applications based on the UDP are significantly less tolerant to packet loss. UDP does not have retransmission facilities. however, retransmissions would almost never help. In an RTP session, by the time a media gateway could receive a retransmission, it would no longer be relative to the reconstructed voice waveform; that part of the waveform in the retransmitted packet would arrive too late.

It is important that bearer and signaling packets are not discarded, otherwise, voice quality or service disruptions might occur. In such instances, CoS mechanisms become very important. By configuring CoS parameters, we can give packets of greater importance a higher priority in the network, thus ensuring packet delivery for critical applications, even during times of network congestion.

Although packet loss of any kind is undesirable, some loss can be tolerated. Some amount of packet loss for voice services could be acceptable as long as the loss is spread out over a large amount of users. As long as the amount of packet loss is less than five percent for the total number of calls, the quality generally is not adversely affected. It is best to drop a packet, versus increasing the latency of all delivered packets by further buffering them.

5.5.5 Reliability

Although network failures are rare, planning for them is essential. Failover strategies are desirable for cases when network devices malfunction or links are broken. An important strategy is to deploy redundant links between network devices and/or to deploy redundant equipment. To ensure continued service, plan carefully for how media gateways and media gateway controllers can make use of the redundant schemes.

IP networks use routing protocols to exchange routing information. As part of their operation, routing protocols monitor the status of interconnecting links. Routing protocols typically detect and reroute packets around a failure if an alternate path exists. Depending on the interconnecting media used for these links, the time taken to detect and recalculate an alternate path can vary. For example, the loss of signal for a SONET/SDH connection can be detected and subsequently rerouted very quickly. However, a connection through an intervening LAN switch might need to time out the keep-alive protocol before a failure is detected.

Having media gateways and media gateway controllers that can actively detect the status of their next-hop address (default gateway) as part of their failover mechanism decreases the likelihood of a large service disruption. Another possible option is that the media gateway and media gateway controller could be directly connected to the router. In this case, the possibility of a link failure (depending on the nature of the failure) could be immediately detected and the network devices would take appropriate action. Still another option for reducing long-term failure could be to employ a redundancy mechanism such as the Virtual Router Redundancy Protocol (VRRP).

5.5.6 Security

Security, especially in a converged voice and data network, is a high priority. We need to protect the voice communications devices from unauthorized access and malicious attack. While we can thwart unauthorized access by using security protocols (such as RADIUS and Ssh), denial-of-service (DoS) attacks can be a real danger to voice services. It is conceivable that such attacks would either cripple or completely disable voice services.

One method of ensuring that DoS attacks are not successful is to use private addressing for the VoIP devices. Private addressing (RFC 1918) can keep Internet-based attacks from happening because private addresses are not routable (not advertised) in the public Internet. (This method would actually only work for most of the Internet and not directly connected autonomous systems because of the possibility of default routing.) We can use private addressing only if the in-band VoIP service never crosses autonomous-system (AS) boundaries. Interface outside the network would be through the ties to the PSTN.

If any part of the VoIP service needs access to the Internet, we can configure packet filtering to provide protection. Through packet filtering, we can selectively allow the VoIP devices to communicate with each other while denying traffic from possible attackers. It is important that whatever packet filtering is employed does not impact the network performance. Any gains made in protecting the equipment from attack could be lost with routers / switches that cannot perform filtering without compromising performance. Security of the actual IP network is also an important consideration. A malicious attack on an IP network, specifically the routers that carry the traffic, could compromise the network services. Someone using a router or similarly capable device connected to the network could spoof the routing protocols and cause disruption.

5.6 Performance Considerations for SS7 over IP

SS7 messages transported over IP networks must meet the stringent performance requirements imposed by both the ITU SS7/C7 standards and user expectations. For example, while the ITU standard specifies that the end-to-end call setup delay cannot exceed 20 to 30 seconds after the ISUP Initial Address Message (IAM) is transmitted; users have generally come to expect much faster response times. For this reason, VoIP networks must be engineered to satisfy user expectations and ITU standards for performance.

5.7 Summary

In this chapter, we introduced the interworking between SS7 and VoIP technology. We defined what VoIP is and the reasons behind the fast evolution of this technology. Then we discussed VoIP functions specially signaling in VoIP networks and how SS7 has an essential role in this kind of network. Finally we viewed the most important considerations in VoIP network which must be taken to ensure a reliable and high quality services.

6. CONCLUSION

The Signaling System 7 (SS7) network plays a crucial role in the functionality of the Public Switched Telephone Network (PSTN). Inter-switch calls would not be completed if Service Switching Points (SSPs) could not exchange signaling messages. Therefore, to ensure consistent and reliable service, the SS7 architecture was developed with built-in redundancy where most important nodes, such as Signal Transfer Points (STPs) and databases, are deployed in pairs or even quads. The SS7 protocol also embeds in its layered structure multiple error checks and procedures to prevent or quickly recover from any link, hardware, or software failure.

In the network that places a high value on reliability, performance is also of great importance. Over years of telephone service availability, users have become accustomed to connection delays of mere seconds. To meet customers' expectations, response time should not increase with increase of signaling traffic resulting from constantly added new services. Traffic keeps increasing because more and more transactions performed in the network require at least one database access. Examples of services that have led to the increase in signaling traffic are Advanced Intelligent Network (AIN) services, Local Number Portability (LNP), etc.

7. REFERENCES

[1] Mamedov F. S., *Telecommunications*, Lecture Notes, Near East University Press, Nicosia, 2000.

[2] Minoli D., *Telecommunication Technology hand Book*, Artech House, Inc., 1991.

[3] Freeman R. L., *Fundamentals of Telecommunications*, John Wiley and Sons, Inc., 2004.

[4] http://www.privateline.com/archive/NTtelephony101.pdf.

[5] http://www.cisco.com/univercd/cc/td/doc.html.

[6] http://www.c7.com/ss7/whitepapers/related/switching_in_ss7.html.

[7] www.c7.com/ss7/whitepapers/ss8_architecture.htm.

[8] Banan M., Signaling System 7 Technology and Applications Overview. http://www.leapforum.org/published/ss7Intro/main.html.

[9] http://www.iec.org/online/tutorials/ss7.

[10] http://www.msforum.org/techinfo/ reports.html.

[11] http://cn.juniper.net/solutions/literature/white_papers/VoIP Protocol/html.