CHAPTER 1

Overview of the PSTN and Comparisons to Voice over IP

The Public Switched Telephone Network (PSTN) has been evolving ever since Alexander Graham Bell made the first voice transmission over wire in 1876. But, before explaining the present state of the PSTN and what's in store for the future, it is important that you understand PSTN history and it's basics. As such, this chapter discusses the beginnings of the PSTN and explains why the PSTN exists in its current state.

This chapter also covers PSTN basics, components, and services to give you a good introduction to how the PSTN operates today. Finally, it discusses where the PSTN could be improved and ways in which it and other voice networks are evolving to the point at which they combine data, video, and voice.

The Beginning of the PSTN

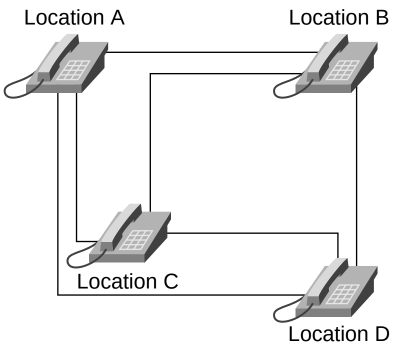
The first voice transmission, sent by Alexander Graham Bell, was accomplished in 1876 through what is called a *ring-down* circuit. A ring-down circuit means that there was no dialing of numbers, Instead, a physical wire connected two devices. Basically, one person picked up the phone and another person was on the other end (no ringing was involved).

Over time, this simple design evolved from a one-way voice transmission, by which only one user could speak, to a bi-directional voice transmission, whereby both users could speak. Moving the voices across the wire required a carbon microphone, a battery, an electromagnet, and an iron diaphragm.

It also required a physical cable between each location that the user wanted to call. The concept of dialing a number to reach a destination, however, did not exist at this time.

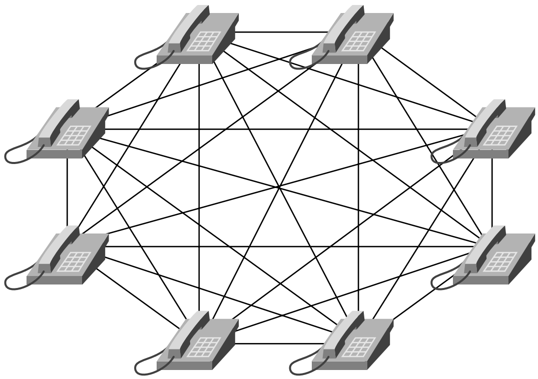
To further illustrate the beginnings of the PSTN, see the basic four-telephone network shown in Figure 1-1. As you can see, a physical cable exists between each location.

Figure 1-1 *Basic Four-Phone Network*



Place a physical cable between every household requiring access to a telephone, however, and you'll see that such a setup is neither cost-effective nor feasible (see Figure 1-2). To determine how many lines you need to your house, think about everyone you call as a value of N and use the following equation: $N \times (N-1)/2$. As such, if you want to call 10 people, you need 45 pairs of lines running into your house.

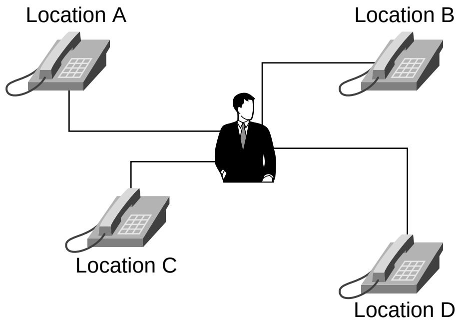
Figure 1-2 *Physical Cable Between All Telephone Users*



Due to the cost concerns and the impossibility of running a physical cable between everyone on Earth who wanted access to a telephone, another mechanism was developed that could map any phone to another phone. With this device, called a *switch*, the telephone users needed only one cable to the centralized switch office, instead of seven.

At first, a telephone operator acted as the switch. This operator asked callers where they wanted to dial and then manually connected the two voice paths. Figure 1-3 shows how the four-phone network example would look today with a centralized operator to switch the calls.

Figure 1-3 *Centralized Operator: The Human Switch*



Now, skip ahead 100 years or so—the human switch is replaced by electronic switches. At this point, you can learn how the modern PSTN network is built.

Understanding PSTN Basics

Although it is difficult to explain every component of the PSTN, this section explains the most important pieces that make the PSTN work. The following sections discuss how your voice is transmitted across a digital network, basic circuit-switching concepts, and why your phone number is 10 digits long.

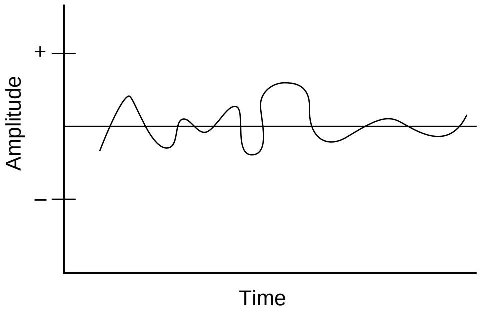
Analog and Digital Signaling

Everything you hear, including human speech, is in analog form. Until several decades ago, the telephony network was based on an analog infrastructure as well.

Although analog communication is ideal for human interaction, it is neither robust nor efficient at recovering from line noise. (*Line noise* is normally caused by the introduction of static into a voice network.) In the early telephony network, analog transmission was passed through amplifiers to boost the signal. But, this practice amplified not just the voice, but the line noise as well. This line noise resulted in an often unusable connection.

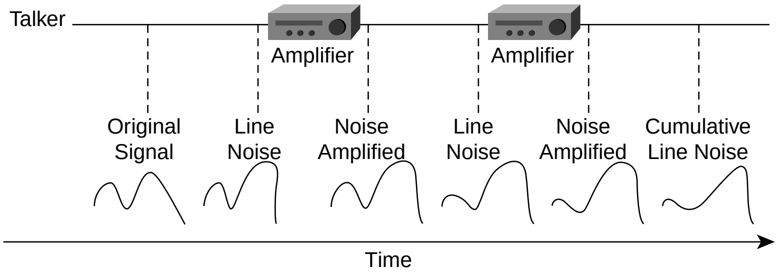
Analog communication is a mix of time and amplitude. Figure 1-4, which takes a high-level view of an analog waveform, shows what your voice looks like through an oscilloscope.

Figure 1-4 *Analog Waveform*



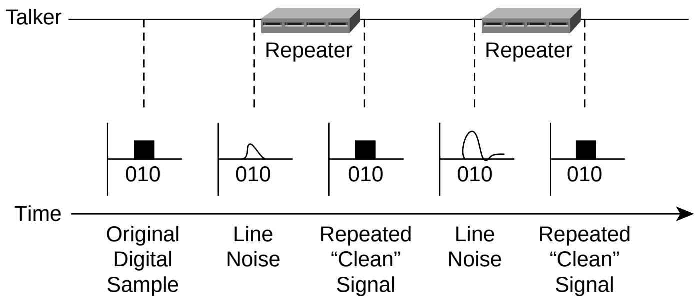
If you were far away from the *end office switch* (which provides the physical cable to your home), an amplifier might be required to boost the analog transmission (your voice). Analog signals that receive line noise can distort the analog waveform and cause garbled reception. This is more obvious to the listener if many amplifiers are located between your home and the end office switch. Figure 1-5 shows that an amplifier does not clean the signal as it amplifies, but simply amplifies the distorted signal. This process of going through several amplifiers with one voice signal is called *accumulated noise*.

Figure 1-5 *Analog Line Distortion*



In digital networks, line noise is less of an issue because repeaters not only amplify the signal, but clean it to its original condition. This is possible with digital communication because such communication is based on 1s and 0s. So, as shown in Figure 1-6, the *repeater* (a digital amplifier) only has to decide whether to regenerate a 1 or a 0.

Figure 1-6 *Digital Line Distortion*



Therefore, when signals are repeated, a clean sound is maintained. When the benefits of this digital representation became evident, the telephony network migrated to *pulse code modulation* (PCM).

Digital Voice Signals

PCM is the most common method of encoding an analog voice signal into a digital stream of 1s and 0s. All sampling techniques use the *Nyquist theorem*, which basically states that if you sample at twice the highest frequency on a voice line, you achieve good-quality voice transmission.

The PCM process is as follows:

- Analog waveforms are put through a voice frequency filter to filter out anything greater than 4000 Hz. These frequencies are filtered to 4000 Hz to limit the amount of crosstalk in the voice network. Using the Nyquist theorem, you need to sample at 8000 samples per second to achieve good-quality voice transmission.
- The filtered analog signal is then sampled at a rate of 8000 times per second.
- After the waveform is sampled, it is converted into a discrete digital form. This sample is represented by a code that indicates the amplitude of the waveform at the instant the sample was taken. The telephony form of PCM uses eight bits for the code and a logarithm compression method that assigns more bits to lower-amplitude signals.

If you multiply the eight-bit words by 8000 times per second, you get 64,000 bits per second (bps). The basis for the telephone infrastructure is 64,000 bps (or 64 kbps).

Two basic variations of 64 kbps PCM are commonly used: μ-law, the standard used in North America; and a-law, the standard used in Europe. The methods are similar in that both use logarithmic compression to achieve from 12 to 13 bits of linear PCM quality in only eight-bit words, but they differ in relatively minor details. The μ-law method has a slight advantage over the a-law method in terms of low-level signal-to-noise ratio performance, for instance.

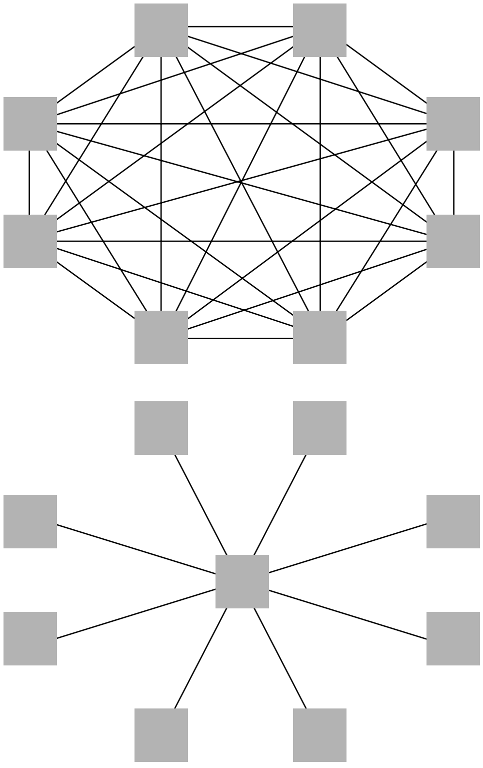
NOTE When making a long-distance call, any µ-law to a-law conversion is the responsibility of the µ-law country.

Local Loops, Trunks, and Interswitch Communication

The telephone infrastructure starts with a simple pair of copper wires running to your home. This physical cabling is known as a *local loop*. The local loop physically connects your home telephone to the central office switch (also known as a *Class 5 switch* or *end office switch*). The communication path between the central office switch and your home is known as the *phone line,* and it normally runs over the local loop.

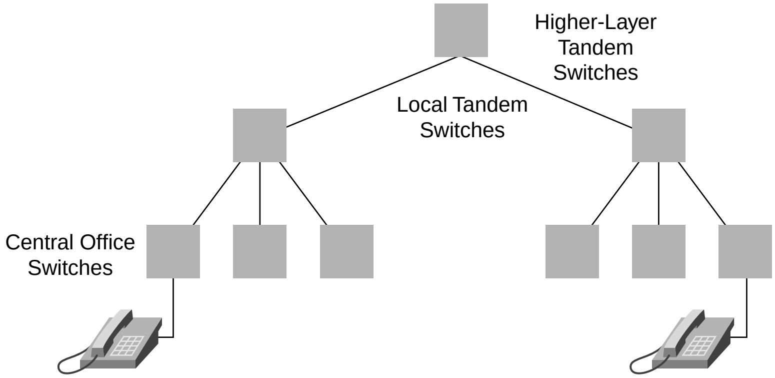
The communication path between several central office switches is known as a *trunk*. Just as it is not cost-effective to place a physical wire between your house and every other house you want to call, it is also not cost-effective to place a physical wire between every central office switch. You can see in Figure 1-7 that a meshed telephone network is not as scalable as one with a hierarchy of switches.

Figure 1-7 *Meshed Network Versus Hierarchical Network*



Switches are currently deployed in hierarchies. End office switches (or central office switches) interconnect through trunks to *tandem switches* (also referred to as Class 4 switches). Higher-layer tandem switches connect local tandem switches. Figure 1-8 shows a typical model of switching hierarchy.

Figure 1-8 *Circuit-Switching Hierarchy*



Central office switches often directly connect to each other. Where the direct connections occur between central office switches depends to a great extent on call patterns. If enough traffic occurs between two central office switches, a dedicated circuit is placed between the two switches to offload those calls from the local tandem switches. Some portions of the PSTN use as many as five levels of switching hierarchy.

Now that you know how and why the PSTN is broken into a hierarchy of switches, you need to understand how they are physically connected, and how the network communicates.

PSTN Signaling

Generally, two types of signaling methods run over various transmission media. The signaling methods are broken into the following groups:

- *User-to-network signaling*—This is how an end user communicates with the PSTN.
- *Network-to-network signaling*—This is generally how the switches in the PSTN intercommunicate.

User-to-Network Signaling

Generally, when using *twisted copper pair* as the transport, a user connects to the PSTN through analog, Integrated Services Digital Network (ISDN), or through a T1 carrier.

The most common signaling method for user-to-network analog communication is *Dual Tone Multi-Frequency (DTMF)*. DTMF is known as in-band signaling because the tones are carried through the voice path. Figure 1-9 shows how DTMF tones are derived.

Figure 1-9 *Dual Tone Multi-Frequency*

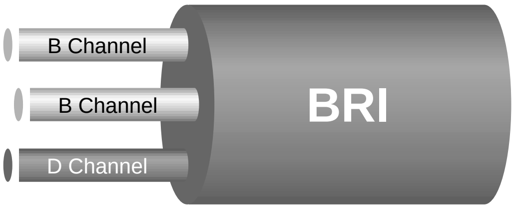
Dual Tone Multi-Frequency

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

When you pick up your telephone handset and press the digits (as shown in Figure 1-9), the tone that passes from your phone to the central office switch to which you are connected tells the switch what number you want to call.

ISDN uses another method of signaling known as *out-of-band*. With this method, the signaling is transported on a channel separate from the voice. The channel on which the voice is carried is called a *bearer* (or B channel) and is 64 kbps. The channel on which the signal is carried is called a data channel (D channel) and is 16 kbps. Figure 1-10 shows a Basic Rate Interface (BRI) that consists of two B channels and one D channel.

Figure 1-10 *Basic Rate Interface*



Out-of-band signaling offers many benefits, including the following:

- Signaling is multiplexed (consolidated) into a common channel.
- Glare is reduced (glare occurs when two people on the same circuit seize opposite ends of that circuit at the same time).
- A lower post dialing delay.
- Additional features, such as higher bandwidth, are realized.

- Because setup messages are not subject to the same line noise as DTMF tones, call completion is greatly increased.

In-band signaling suffers from a few problems, the largest of which is the possibility for *lost tones*. This occurs when signaling is carried across the voice path and it is a common reason why you can sometimes experience problems remotely accessing your voice mail.

Network-to-Network Signaling

Network-to-network communication is normally carried across the following transmission media:

- T1/E1 carrier over twisted pair

  T1 is a 1.544-Mbps digital transmission link normally used in North America and Japan.

  E1 is a 2.048-Mbps digital transmission link normally used in Europe.

- T3/E3, T4 carrier over coaxial cable

  T3 carries 28 T1s or 672 64-kbps connections and is 44.736 Mbps.

  E3 carries 16 E1s or 512 64-kbps connections and is 34.368 Mbps.

  T4 handles 168 T1 circuits or 4032 4-kbps connections and is 274.176 Mbps.

- T3, T4 carrier over a microwave link
- Synchronous Optical Network (SONET) across fiber media

  SONET is normally deployed in OC-3, OC-12, and OC-48 rates, which are 155.52 Mbps, 622.08 Mbps, and 2.488 Gbps, respectively.

Network-to-network signaling types include in-band signaling methods such as Multi-Frequency (MF) and Robbed Bit Signaling (RBS). These signaling types can also be used to network signaling methods.

Digital carrier systems (T1, T3) use A and B bits to indicate on/off hook supervision. The A/B bits are set to emulate Single Frequency (SF) tones (SF typically uses the presence or absence of a signal to signal A/B bit transitions). These bits might be *robbed* from the information channel or multiplexed in a common channel (the latter occurs mainly in Europe). More information on these signaling types is found in Chapter 3, “Basic Telephony Signaling.”

MF is similar to DTMF, but it utilizes a different set of frequencies. As with DTMF, MF tones are sent in-band. But, instead of signaling from a home to an end office switch, MF signals from switch to switch.

Network-to-network signaling also uses an out-of-band signaling method known as *Signaling System 7* (SS7) (or C7 in European countries). This section covers some of the benefits of SS7, however SS7 is covered in depth in Chapter 4, “Signaling System 7.”

NOTE SS7 is beneficial because it is an out-of-band signaling method and it interconnects to the Intelligent Network (IN). Connection to the IN enables the PSTN to offer Custom Local Area Signaling Services (CLASS) services.

SS7 is a method of sending messages between switches for basic call control and for CLASS. These CLASS services still rely on the end-office switches and the SS7 network. SS7 is also used to connect switches and databases for network-based services (for example, 800-number services and Local Number Portability [LNP]).

Some of the benefits of moving to an SS7 network are as follows:

- Reduced post-dialing delay

  There is no need to transmit DTMF tones on each hop of the PSTN. The SS7 network transmits all the digits in an initial setup message that includes the entire calling and called number. When using in-band signaling, each MF tone normally takes 50 ms to transmit. This means you have at least a .5-second post-dialing delay per PSTN hop. This number is based on 11-digit dialing (11 MF tones $\times$ 50 ms = 550 ms).

- Increased call completion

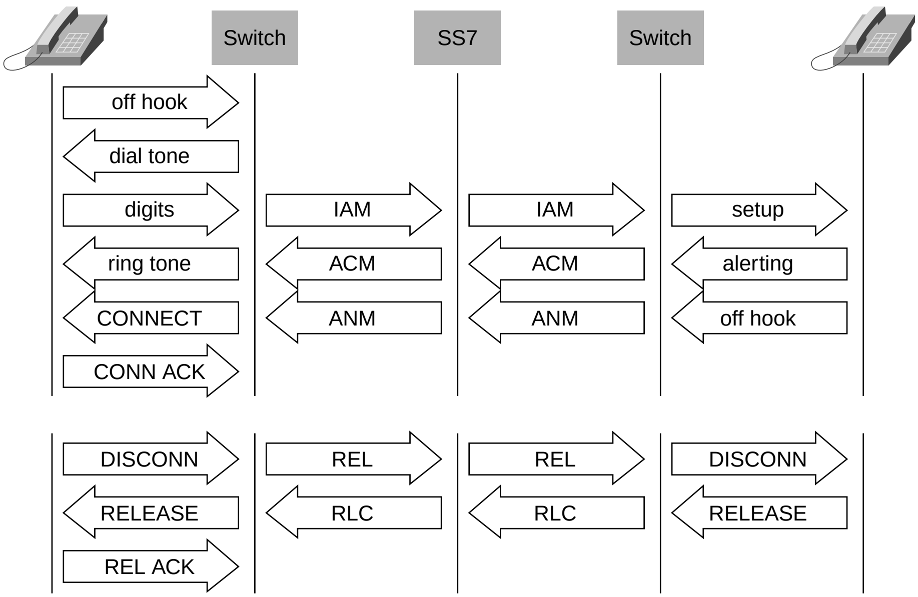
  SS7 is a packet-based, out-of-band signaling protocol, compared to the DTMF or MF in-band signaling types. Single packets containing all the necessary information (phone numbers, services, and so on) are transmitted faster than tones generated one at a time across an in-band network.

- Connection to the IN

  This connection provides new applications and services transparently across multiple vendors' switching equipment as well as the capability to create new services and applications more quickly.

To further explain the PSTN, visualize a call from my house to my Grandma's house 10 miles away. This call traverses an end office switch, the SS7 network (signaling only), and a second end office switch. Figure 1-11 displays the call flow from my house to Grandma's.

Figure 1-11 *PSTN Call Flow to Grandma's House*



To better explain the diagram in Figure 1-11, let's walk through the flow of the call:

1. I pick up the phone and send an off-hook indication to the end office switch.
2. The switch sends back a dial tone.
3. I dial the digits to call Grandma's house (they are sent in-band through DTMF).
4. The switch interprets the digits and sends an Initial Address Message (IAM, or setup message) to the SS7 network.
5. The SS7 network reads the incoming IAM and sends a new IAM to Grandma's switch.
6. Grandma's switch sends a setup message to Grandma's phone (it rings her phone).
7. An alerting message (alerting is the same as the phone ringing) is sent from Grandma's switch (not from her phone) back to the SS7 network through an Address Complete Message (ACM).
8. The SS7 network reads the incoming ACM and generates an ACM to my switch.
9. I can hear a ringing sound and know that Grandma's phone is ringing. (The ringing is not synchronized; your local switch normally generates the ringing when the ACM is received from the SS7 network.)
10. Grandma picks up her phone, sending an off-hook indication to her switch.

11 Grandma's switch sends an ANswer Message (ANM) that is read by the SS7, and a new ANM is generated to my switch.

12 A connect message is sent to my phone (only if it's an ISDN phone) and a connect acknowledgment is sent back (again, only if it's an ISDN phone). (If it is not an ISDN phone, then on-hook or off-hook representations signal the end office switch.)

13 I can now talk to Grandma until I hang up the phone (on-hook indication).

If Grandma's phone was busy, I could use an IN feature by which I could park on her line and have the PSTN call me back after she got off the phone.

Now that you have a basic understanding of how the PSTN functions, the next section discusses services and applications that are common in the PSTN.

If you want more information on PSTN signaling types, see Chapter 3 and Chapter 4.

PSTN Services and Applications

As with almost every industry, it is usually better and easier to acquire additional business from current customers than it is to go out and get new customers. The PSTN is no different. Local Exchange Carriers (LECs) have been increasing the features they offer to create a higher revenue stream per consumer.

Numerous services are now available, for example, which were not available just a few years ago. These services come in two common flavors: *custom calling* features and CLASS features.

Custom calling features rely upon the end office switch, not the entire PSTN, to carry information from circuit-switch to circuit-switch. CLASS features, however, require SS7 connectivity to carry these features from end to end in the PSTN.

The following list includes a few of the popular custom calling features commonly found in the PSTN today:

- Call waiting—Notifies customers who already placed a call that they are receiving an incoming call.
- Call forwarding—Enables a subscriber to forward incoming calls to a different destination.
- Three-way calling—Enables conference calling.

With the deployment of the SS7 network, advanced features can now be carried end to end. A few of the CLASS features are mentioned in the following list:

- Display—Displays the calling party's directory number, or Automatic Number Identification (ANI).

- Call blocking—Blocks specific incoming numbers so that callers are greeted with a message saying the call is not accepted.
- Calling line ID blocking—Blocks the outgoing directory number from being shown on someone else's display. (This does not work when calling 800-numbers or certain other numbers.)
- Automatic callback—Enables you to put a hold on the last number dialed if a busy signal is received, and then place the call after the line is free.
- Call return (*69)—Enables users to quickly reply to missed calls.

A majority of these features are possible due to the use of SS7 and the IN. Many interexchange carriers (IXCs) also offer business features, such as the following:

- Circuit-switched long distance—Basic long-distance services (normally at a steeply discounted rate).
- Calling cards—Pre-paid and post-paid calling cards. You dial a number, enter a password, and then call your destination.
- 800/888/877 numbers—The calling party is not charged for the call; Rather, the party called is charged (normally at a premium rate).
- Virtual Private Networks (VPNs)—The telephone company manages a private dialing plan. This can greatly reduce the number of internal Information Service (IS) telecommunications personnel.
- Private leased lines—Private leased lines from 56 kbps to OC-48s enable both data and voice to traverse different networks. The most popular speed by far in North America is T1.
- Virtual circuits (Frame Relay or Asynchronous Transfer Mode [ATM])—The telephone carrier (IXC or LEC) switches your packets. It does this packet by packet (or cell by cell in ATM), not based upon a dedicated circuit.

This list of IXC business features is merely a sampling of the more popular features and applications available in the PSTN. Although the PSTN is evolving and consumers are using more of its features, the basic user experience has remained somewhat consistent since the inception of digital networking for telephony communications.

PSTN Numbering Plans

One feature that slowly changed over time is the dial plan. The addition of second lines for Internet access, cell phones, and fax machines has created a relative shortage of phone numbers. The next section delves into how the PSTN dial plan is put together and what you can expect over the next few years.

In some places in the United States, it is necessary to dial 1+10 digits for even a local call. This will become more and more prevalent as more devices require telephone numbers. The need to dial 1+10 digits for a local number is normally due to an *overlay*. An overlay can result in next-door neighbors having different area codes. An overlay is when a region with an existing area code has another area code "overlayed." This offers the existing customers the benefits of not having to switch area codes, but forces everyone in that region to dial 10 digits to call anywhere.

Essentially, two numbering plans are used with the PSTN: the North American Numbering Plan (NANP) and the International Telecommunication Union Telecommunication Standardization Sector (ITU-T; formerly CCITT) International Numbering Plan. They are discussed in the following sections.

NANP

NANP is an 11-digit dialing plan that contains three parts: the Numbering Plan Area (NPA, also referring to as area code), Central Office Code (NXX), and Station Number. This plan is often referred to as NPA-NXX-XXXX.

NPA codes use the following format:

> NXX, where N is a value between 2–9 and X is a value between 0–9.

NANP is also referred to as 1+10. This means that when a 1 is the first number dialed, it will be proceeded by a 10-digit NPA-NXX-XXXX number. This enables the end office switch to determine whether it should expect a 7- or 10-digit telephone number.

Your LEC keeps track of what long-distance provider you use in a static table on the end office switch. Each long-distance carrier has a code. This long-distance code is assigned by the North American Numbering Plan Association (NANPA) and is added to the number you call so that it is routed to the proper long-distance network carrier (or IXC).

NOTE Popular today, carrier-selection numbers are used to have a "secondary" long-distance carrier. Dial-around numbers allow you to choose a long-distance carrier call by call by adding 7 digits to each outgoing call. Much advertising has been done to have telephony users specify 10+XX+XXX to not use their primary carrier.

The reason for carrier selection is simple. You don't have to switch and can use different LD carriers based upon the time of day, week, location called, type of call, or personal preference.

ITU-T International Numbering Plan

ITU-T Recommendation E.164 specifies that a Country Code (CC), National Destination Code (NDC), and Subscriber Number (SN) be used to route a call to a specific subscriber.

The CC consists of one, two, or three digits. The first digit (1–9) defines world numbering zones. A list of all the defined CCs is found in ITU-T Recommendation E.164 Annex A.

NDC and SN vary in length based on the needs of the country. Neither one has more than 15 digits.

Many other recommendations and specifications for international number plans are found in the E. recommendations from the ITU-T.

Although dial plans might not seem extremely important at the moment, they are crucial to the successful deployment and implementation of Voice over IP (VoIP) or traditional circuit-switched networks.

Regardless of which dialing plan is used in your country, you can expect to see changes in the ways you can dial as well as whom you dial.

Drivers Behind the Convergence Between Voice and Data Networking

Understanding PSTN basics includes knowing why the existing PSTN does not fit all the needs of its builders or users. After you understand where today's PSTN is lacking, you will know where to look to find a solution. This section sets the stage for why the voice and data networks are merging into a signal network.

Drawbacks to the PSTN

Although the PSTN is effective and does a good job at what it was built to do (that is, switch voice calls), many business drivers are striving to change it to a new network, whereby voice is an application on top of a data network. This is happening for several reasons:

- Data has overtaken voice as the primary traffic on many networks built for voice.

  Data is now running on top of networks that were built to carry voice efficiently. Data has different characteristics, however, such as a variable use of bandwidth and a need for higher bandwidth.

  Soon, voice networks will run on top of networks built with a data-centric approach. Traffic will then be differentiated based upon application instead of physical circuits. New technologies (such as Fast Ethernet, Gigabit Ethernet, and Optical Networking) will be used to deploy the high-speed networks that needed to carry all this additional data.

- The PSTN cannot create and deploy features quickly enough.

With increased competition due to deregulation in many telecommunications markets, LECs are looking for ways to keep their existing clientele. The primary method of keeping customers is by enticing them through new services and applications.

The PSTN is built on an infrastructure whereby only the vendors of the equipment develop the applications for that equipment. This means you have one-stop shopping for all your needs. It is very difficult for one company to meet all the needs of a customer. A more open infrastructure, by which many vendors can provide applications, enables more creative solutions and applications to be developed. It is also not possible with the current architecture to enable many vendors to write new applications for the PSTN. Imagine where the world would be today if vendors, such as Microsoft, did not want other vendors to write applications for its software.

- Data/Voice/Video (D/V/V) cannot converge on the PSTN as currently built.

  With only an analog line to most homes, you cannot have data access (Internet access), phone access, and video access across one 56-kbps modem. High-speed broadband access, such as digital subscriber line (DSL), cable, or wireless, is needed to enable this convergence. After the last bandwidth issues are resolved, the convergence can happen to the home. In the backbone of the PSTN, the convergence has already started.

- The architecture built for voice is not flexible enough to carry data.

  Because the bearer channels (B channels and T1 circuits), call-control (SS7 and Q.931), and service logic (applications) are tightly bound in one closed platform, it is not possible to make minor changes that might improve audio quality.

It is also important to note that circuit-switched calls require a permanent 64-kbps dedicated circuit between the two telephones. Whether the caller or the person called is talking, the 64-kbps connection cannot be used by any other party. This means that the telephone company cannot use this bandwidth for any other purpose and must bill the parties for consuming its resources.

Data networking, on the other hand, has the capability to use bandwidth only when it is required. This difference, although seemingly small, is a major benefit of packet-based voice networking.

Telecommunications Deregulation

So far, you have looked at the technical issues of how the PSTN operates, the basic hierarchy, and why you might need to converge voice and data networks. One important reason for this convergence is more political than technical.

Various countries throughout Europe, Asia, and the Americas are opening up their telecommunications markets to competition. In addition, in some cases, they are selling off the existing government-run telephone carriers to a private company (or many companies).

In the United States, a publicly owned utility ran the PSTN from its inception until its divestiture in the early 1980s. In many other countries, however, the government ran the PSTN. This is changing as governments realize that communication is important to survival in the next century. These governments also realize that with communication comes knowledge, and with knowledge comes strength and prosperity.

Many new voice carriers are rushing to join these new deregulated markets. With the influx of fresh competition, pricing models are changing, and new, as well as old, carriers are considering deploying the latest technology to lower the cost of doing business.

The additional advantage of deploying new technology is the ability to offer value-added services and deploy these new services in a short amount of time. Services include bundled voice and Internet access, unified communications, Internet call waiting, and others.

Let’s use the United States as an example of how competition affects the telecommunications marketplace by taking a look at the breakup of the utility in the early 1980s. American Telephone and Telegraph (AT&T) signed a divestiture agreement that stated it would divest itself of its 22 telephone operating companies. These 22 telephone companies were placed into 7 holding companies, which came to be known as the LECs.

AT&T was broken into a long-distance carrier or an IXC, which kept the AT&T name, and many regional Bell operating companies (RBOCs). These RBOCs actually provided the local loop and line to everyone in their local regions.

The U.S. RBOCs (Pacific Telesis, Southwestern Bell, Nynex, Bell Atlantic, Southern Bell, US West, and NYNEX) all had areas known as Local Area and Transport Areas (LATAs), which are local calling areas. These RBOCs were also known as LECs. If these LECs wanted to pass traffic between LATAs, they had to use an IXC.

As a result, many IXCs (AT&T, MCI, Sprint, and others) could offer long-distance domestic service and develop agreements with international carriers to provide international services. The local LECs, however, were not allowed to provide long-distance service, and pricing was highly regulated to avoid monopolies.

When competition arose in the LEC market, the existing LECs were then called Incumbent LECs (ILECs) and the newcomers were called Competitive LECs (CLECs).

Many of the ILECs have started to consolidate. They are currently attempting to meet certain requirements to be able to enter the long-distance marketplace. This will enable them to bypass such IXCs as AT&T and MCI and keep the money they normally pay them for long-distance service.

More recently, new competitors to LECs, CLECs, and IXCs have emerged. These competitors come in the form of Internet telephony service providers (ITSPs) and Greenfield carriers. ITSPs are Internet service providers (ISPs) that add voice functionality to their portfolio and carry voice traffic across data networks, which frequently span traditional ILEC boundaries.

Greenfield carriers are carriers that build networks from scratch (for instance, a data network built specifically to carry packet voice) instead of using circuit-switching networks normally used by the LECs and IXCs.

NOTE Although deregulation and competition have existed in the U.S. since 1982, their emergence in other countries is more recent. China and Germany, for example, have opened up their respective telephone markets only within the last few years.

In different countries' respective markets, these competitors can exploit different PSTN market niches. A Greenfield carrier in any country might attempt to offer business customers both voice and data over a high-speed data infrastructure. Meanwhile, a traditional CLEC might attempt to offer high-speed access to both residential customers and businesses over traditional access, such as T1 circuits, as well as new high-bandwidth services, such as DSL.

Packet Telephony Network Drivers

The previous section discussed political drivers for competition in the PSTN. This section explains why a carrier might choose to develop a packet telephony network in lieu of a traditional circuit-switching network.

The integration of D/V/V is more than just a change in infrastructure. D/V/V integration also enables new features to be developed more quickly and opens up application development to thousands of Independent Software Vendors (ISVs). You can compare this integration of D/V/V to the change from mainframe computers, for which very few vendors developed applications, to client/servers, for which multiple vendors developed applications for distributed systems.

Figure 1-12 shows how the circuit-switching model is breaking into a new model by which open standards exist between all three layers. A packet infrastructure will carry the actual voice (media), the call-control layer will be separate from the media layer, and open APIs (Application Programming Interfaces) will enable new services to be created by ISVs.

Figure 1-12 *Circuit Switching Versus Packet Switching*

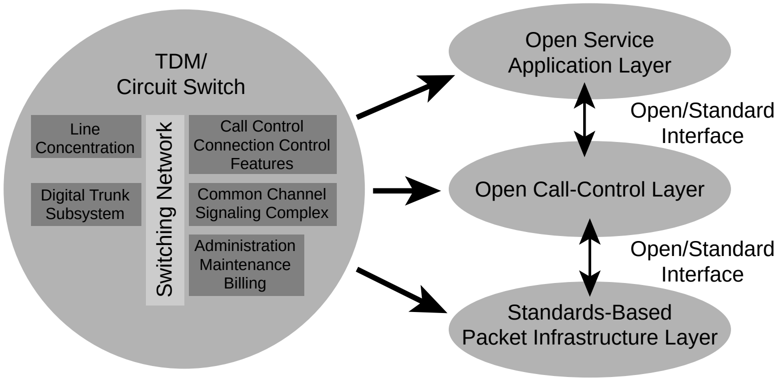


Figure 1-12 is an over-simplification of the changes that are actually happening. To further discuss these changes, you need to take a closer look at each of the three layers.

Standards-Based Packet Infrastructure Layer

The packet infrastructure replaces the circuit-switching infrastructure in this new model. This infrastructure most likely will be IP, although this model also works if ATM is the underlying transport and IP rides across the top. IP is so attractive as the packet infrastructure because of its ubiquitous nature and the fact that it is the de facto application interface. This means that software applications running over IP do not have to be known. IP simply transports the data end to end, with no real interest in the payload.

NOTE To provide the proper prioritization on a *congested* IP network, the IP network must have some knowledge of the applications.

Real-time Transport Protocol (RTP) is utilized in addition to a User Datagram Protocol (UDP)/IP header to provide *timestamping*. RTP runs atop UDP and IP and is commonly noted as RTP/UDP/IP. RTP is currently the cornerstone for carrying real-time traffic across IP networks. (Microsoft Netmeeting, for instance, utilizes RTP to carry audio and video communications.) To date, all VoIP signaling protocols utilize RTP/UDP/IP as their transport mechanism for voice traffic. Often, RTP packet flows are known as *RTP streams*. This nomenclature is used to describe the audio path.

In IP networks, it is common and normal for packet loss to occur. In fact, Transmission Control Protocol/Internet Protocol (TCP/IP) was built to utilize packet loss as a means of controlling the flow of packets. In TCP/IP, if a packet is lost, it is retransmitted. In most real-time applications, retransmission of a packet is worse than receiving a packet due to the time-sensitive nature of the information.

The ITU-T recommends a one-way delay of no more than 150 ms. In a Cisco VoIP network, the unidirectional delay might be 120 ms (currently, 65 ms to 85 ms of that 120-ms delay is derived from two Cisco VoIP gateways when using G.729). If the receiving station must request that a packet be re-transmitted, the delay will be too large, and large gaps and breaks in the conversation will occur.

NOTE The RTP stream is also often referred to as the *media stream*. Therefore, you can use IP in conjunction with UDP and RTP to replace a traditional 64-kbps voice circuit.

RTP has a field that stamps the exact time the packet was sent (in relation to the entire RTP stream). This information is known as *RTP timestamps* and is used by the device terminating/receiving the audio flow. The receiving device uses the RTP timestamps to determine when a packet was expected, whether the packet was in order, and whether it was received when expected. All this information helps the receiving station determine how to tune its own settings to mask any potential network problems such as delay, jitter, and packet loss.

NOTE *Jitter* is the variation of interpacket arrival time, or the difference between when a packet is supposed to be received and when it is actually received.

One of the main benefits of IP is the fact that properly built IP networks are *self-healing*. This means that because dynamic routing protocols are used and multiple possible destinations exist, a network can re-converge based upon the best route. It also means that it is possible for your voice (packetized in IP) to take multiple paths to the same destination. Currently you cannot nail up a single path between two destinations. Each individual packet takes the best path between sender and receiver.

The fact that the packet layer is based upon open standards enables multiple vendors to provide solutions that are interoperable. These open standards enable increased competition at this packet layer. The ITU-T, Internet Engineering Task Force (IETF), European Telecommunication Standards Institute (ETSI), and EIA-TIA are only a few of the standards bodies you might be familiar with.

One key component of having a standards-based packet infrastructure is the ability to have open standards to layers at the call-control layer. Referring to Figure 1-12, these open standards are provided by protocols such as H.323, SGCP, MGCP, SIP, and so on, which have open defined interfaces and are widely deployed into the packet infrastructure. One of the jobs of the call-control protocol is to tell the RTP streams where to terminate and where to begin. Call-control accomplishes this task by translating between IP addresses and phone numbering plans.

Open Call-Control Layer

Call-control, in a nutshell, is the process of making a routing decision about where a call needs to go and somehow making the call happen. In the PSTN today, these decisions are carried out by SS7 and are made by Service Control Points (SCPs). Chapter 8, "VoIP: An In-Depth Analysis," discusses how the different VoIP protocols work and how they solve different network design issues.

In this new model of separating the bearers (RTP streams) from the call-control layer and separating the call-control layer from the services, it is necessary to make sure that standards-based protocols are used. Data networking is unique in the fact that multiple protocols can co-exist in a network and you can tailor them to the particular needs of the network.

Many different IP routing protocols exist, for example, and each is specifically designed for a certain type of network. Some of these include the Router Information Protocol (RIP), Interior Gateway Routing Protocol (IGRP), Enhanced Interior Gateway Routing Protocol (EIGRP), Intermediary System to Intermediary System (IS-IS), Open Shortest Path First (OSPF), and Border Gateway Protocol (BGP). Each protocol solves a similar problem—routing updates. Each routing problem is slightly different, however, and requires a different tool. In this case, the tool is a routing protocol, which solves each problem.

You can say the same of VoIP call-control protocols. They all solve a similar problem—phone numbering to IP address translation; however, they might all be used for slightly different purposes.

For instance, currently H.323 is the most widely deployed VoIP call-control protocol. H.323, however, is not widely seen as a protocol that is robust enough for PSTN networks. For these networks, other protocols such as Media Gateway Control Protocol (MGCP) and Session Initiation Protocol (SIP) are being developed.

Many VoIP call-control protocols are being developed, so it is possible that different protocols will be deployed throughout the coming years. Each protocol will be developed to fix a certain problem and serve a particular purpose. A leader will emerge from the mud, but only if there must be a winner. For the short term, at least, many protocols will be used, and there will be no need for a single call-control protocol.

VoIP Call-Control Protocols

As of this writing, the main VoIP call-control protocols are H.323, Simple Gateway Control Protocol (SGCP), Internet Protocol Device Control (IPDC), MGCP, and SIP. They are defined as follows:

- H.323 is the ITU-T recommendation with the largest installed base, simply because it has been around the longest and no other protocol choices existed before H.323. Chapter 10, "H.323," discusses this protocol in detail.
- SGCP was developed starting in 1998 to reduce the cost of endpoints (gateways) by having the intelligent call-control occur in a centralized platform (or gateway controller). Chapter 12, "Gateway Control Protocols," covers this in more detail.
- IPDC is very similar to SGCP, but it has many other mechanisms for operations, administration, management, and provisioning (OAM&P) than SGCP. OAM&P is crucial to carrier networks because it covers how they are maintained and deployed.
- In late 1998, the IETF put IPDC and SGCP in a room and out popped MGCP. MGCP is basically SGCP with a few additions for OAM&P. MGCP is covered in more detail in Chapter 12.
- SIP is being developed as a media-based protocol that will enable end devices (endpoints or gateways) to be more intelligent, and enable enhanced services down at the call-control layer. Chapter 11, "Session Initiation Protocol," covers SIP in detail.

To briefly explain the various differences between these call-control protocols, let's take a look at how they signal endpoints.

H.323

H.323 is an ITU-T recommendation that specifies how multimedia traffic is carried over packet networks. H.323 utilizes existing standards (Q.931, for example) to accomplish its goals. H.323 is a rather complex protocol that was not created for simple development of applications. Rather, it was created to enable multimedia applications to run over "unreliable" data networks. Voice traffic is only one of the applications for H.323. Most of the initial work in this area focused on multimedia applications, with video and data-sharing a major part of the protocol.

Applications require significant work if they are to be scalable with H.323; for example, to accomplish a call transfer requires a separate specification (H.450.2). SGCP and MGCP, on the other hand, can accomplish a call transfer with a simple command, known as a modify connection (MDCX), to the gateway or endpoint. This simple example represents the different approaches built into the protocol design itself—one tailored to large deployment for simple applications (MGCP), and the other tailored to more complicated applications but showing limitations in its scalability (H.323).

To further demonstrate the complexity of H.323, Figure 1-13 shows a call-flow between two H.323 endpoints.

Figure 1-13 *H.323 Call-Flow*

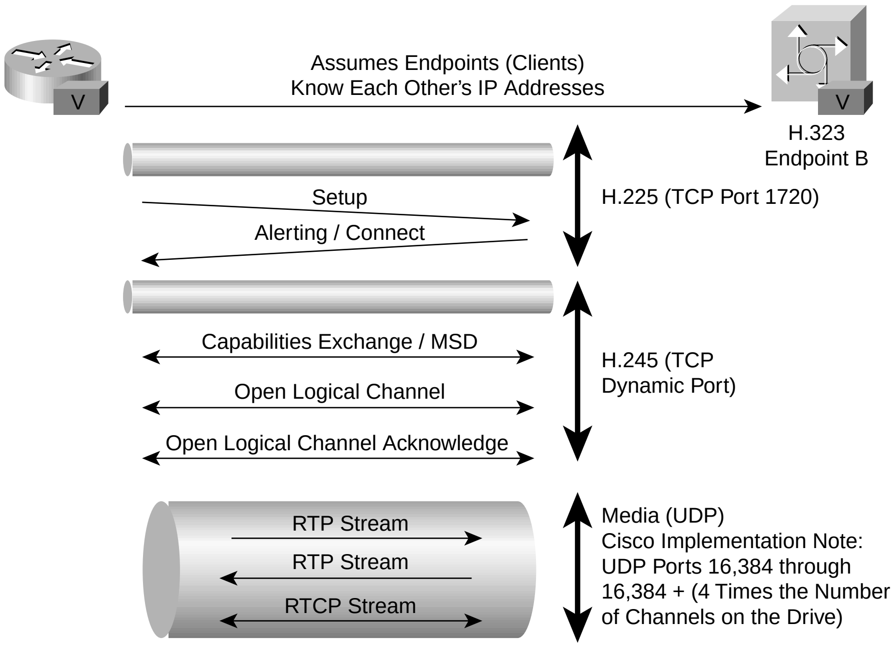


Figure 1-13 illustrates the most basic H.323 call-flow. In most cases, more steps are needed because gatekeepers are involved.

To better explain Figure 1-13, let's step through the call-flow:

1. Endpoint A sends a setup message to Endpoint B on TCP Port 1720.
2. Endpoint B replies to the setup message with an alerting message and a port number to start H.245 negotiation.
3. H.245 negotiation includes codec types (G.729 and G.723.1), port numbers for the RTP streams, and notification of other capabilities the endpoints have.
4. Logical channels for the UDP stream are then negotiated, opened, and acknowledged.
5. Voice is then carried over RTP streams.
6. Real Time Transport Control Protocol is used to transmit information about the RTP stream to both endpoints.

This call-flow is based on H.323 v1. H.323 v2, however, enables H.245 negotiation to be tunneled in the H.225 setup message. This is known as *fast-start*, and it cuts down on the number of roundtrips required to set up an H.323 call. It does not, however, make the protocol any less complex. More detailed analysis of H.323 is found in Chapter 10.

SGCP and MGCP

SGCP and MGCP were developed to enable a central device, known as a Media Gateway Controller (MGC) or *soft-switch*, to control endpoints or Media Gateways (MGs). Both of those protocols are referenced simultaneously as *xGCP*. You can develop applications through the use of standard-based APIs that interface with the MGCs and offer additional functionality (such as call waiting and CLASS features) and applications.

The Cisco version of this technology is known as the Virtual Switch Controller (VSC). In this scenario, the entire IP network acts like one large virtual switch, with the VSC controlling all the MGs.

Figure 1-14 shows how a typical network design works with a virtual switch running MGCP.

Figure 1-14 *Virtual Switch Controller*

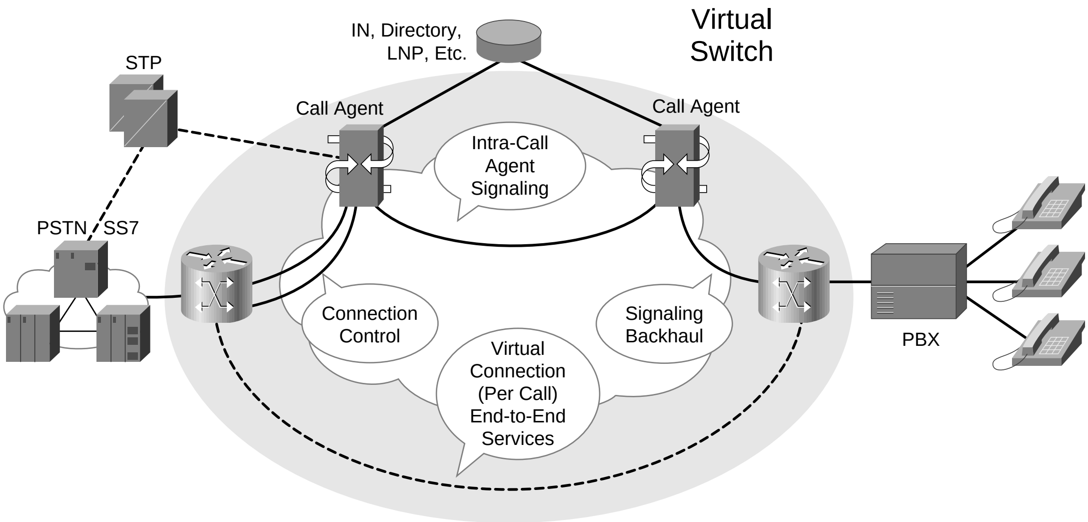


Figure 1-14 also shows how the legacy PSTN and enterprise networks are connected to gateways or endpoints that enable access into the new packet network. This packet gateway receives direction from the Call Agent (VSC), which can communicate with the SS7 network and the IN and can tell the gateways or endpoints how and when to set up the call.

To understand Figure 1-14 in greater detail, all the various components must be described. The existing PSTN/SS7 network is connected to the Switching Transfer Point (STP), which also is connected to the MGC or Call Agent. This connection is where the signaling (SS7) takes place.

The PSTN/SS7 network is also connected to an MG, which is a signal-less trunk that is often known as an *Inter-Machine Trunk* or IMT. The MG is where the 64-kbps voice trunks are converted into packets and placed onto the IP network.

The MGCs or Call Agents also intercommunicate. This protocol is currently undefined in the standards bodies. Based on the current state of the industry, however, it appears that a variant of SIP or ISDN User Part (ISUP) over IP—a portion of SS7 running on top of IP—will be the primary protocol. The MGCs have a connection to the IN (described earlier in this chapter) to provide CLASS services. The MGCs receive signals from the SS7 network and tell the MGs when to set up IP connections and with which other MGs they should set them up.

The MG on the right side of Figure 1-14 does not have a connection to the SS7 network. Therefore, a mechanism known as *signaling backhaul* must be used to tell the VSC when and how a call is arriving. Signaling backhaul is normally done with ISDN. The MG or some other device separates the D channel from the B channels and forwards the D channel to the MGC through IP. Signaling backhaul is currently undefined in the standards bodies. By the time this book is printed, however, there should be a specification for signaling backhaul.

For a more detailed explanation of how all these components work together, see Chapter 12, "Gateway Control Protocols" and Chapter 13, "Virtual Switch Controller."

SIP

SIP is best described by RFC 2543, which states that it is an application-layer control (signaling) protocol for creating, modifying, and terminating sessions with one or more participants.

These multimedia sessions include audio, video, and data and can include multiple partners. SIP enables participants to be invited into an impromptu conference. These multimedia sessions can communicate through multicast, unicast, or a combination of both delivery mechanisms.

Very few implementations of SIP are currently running, although many vendors and customers are interested in using SIP to deploy enhanced services.

See Chapter 11 for more detailed information on SIP.

Open Service Application Layer

By far the most interesting layer of any networking protocol is the application layer. Without good applications, the network infrastructure is built for naught. When moving to a new infrastructure, it is not necessary to carry over all the features that are on the old infrastructure. Only the features or applications that customers need are required.

When building a network that has open interfaces from the packet layer to the call-control layer and from the call-control layer to the application layer, vendors no longer have to develop applications. Now, they can simply write to these standard APIs and have access to a whole new infrastructure. When a new packet infrastructure is built, opportunities for new applications become widely available.

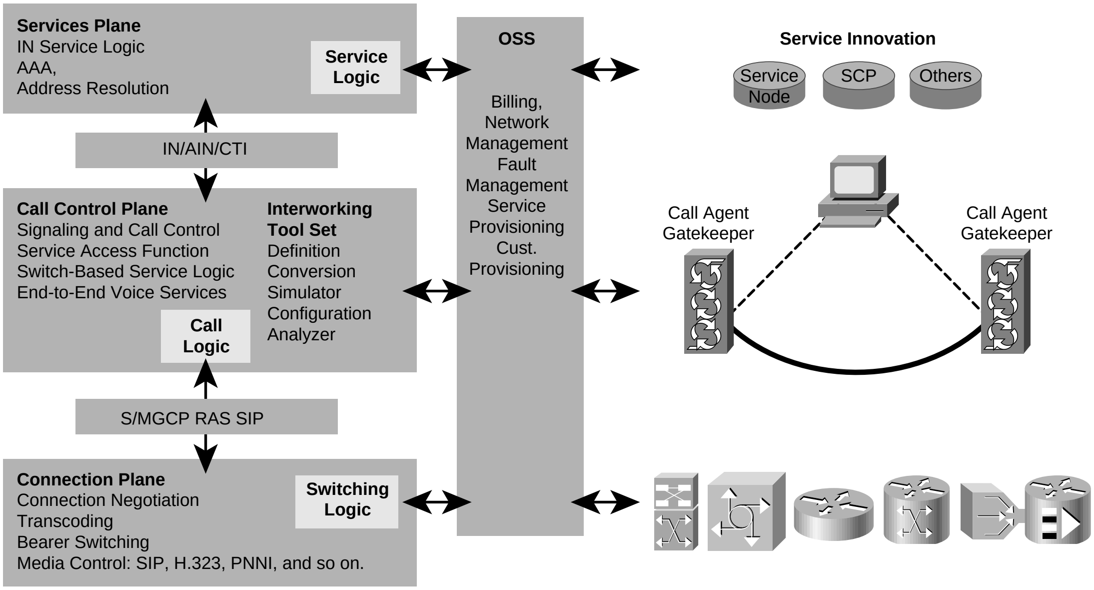
Legacy applications such as call-centers for enterprise networks, and standard PSTN applications such as call waiting and call forwarding, must be ported onto a new infrastructure without the end user realizing that the change occurred. After these legacy applications are ported, literally thousands of new enhanced applications can be specifically developed for packet infrastructures. These include (but are not limited to) Internet call waiting, push to talk, find me-follow me, and unified messaging. These applications are discussed in Chapter 6, "Voice over IP Benefits and Applications."

New PSTN Network Infrastructure Model

As discussed in the previous sections, the new infrastructure will focus on the ability to separate the old stagnant infrastructure into a model by which multiple vendors can develop applications and features quickly for the consumer. Figure 1-15 shows how Cisco Systems wants to carry this model forward.

Figure 1-15 clearly shows the relationship between all three layers as well as the relationship between these layers and the components that would be used in a live network. Carriers will enjoy this method, as it means they won't be locked into a single solution for any of their layers. They will be able to mix and match all three layers to offer the services, functionality, and time-to-market that they need.

Figure 1-15 *Elements of Packet Telephony*



Some carriers might be hesitant to utilize more than one equipment vendor to cut down on their integration timeframe, but many service providers will partner with a minimum of two vendors to ensure competition.

The reality of Figure 1-15 is that the bearers, connection plane, or media transport will be either IP gateways or ATM gateways, or a combination of both. Multiple vendors will be in this space initially, but most likely, they will consolidate to three to five major players.

NOTE A common trend in the manufacturing and carrier arena is *consolidation*. The consolidation of manufacturers is one reason for the dramatic reduction in the number of players in this space.

The call-control plane is an extremely important piece of the new PSTN network infrastructure model, as it must gracefully coexist with both the connection plane and the service (application) layers. Many vendors are building MGC technology.

In fact, the authors are working with approximately 15 vendors to ensure compatibility from the connection plane into the call-control and service/application plane.

Many vendors will continue to be in the call-control plane, as service providers will more than likely use several vendors for this key technology, depending upon what service they decide to deploy. The onus on the Call Agent vendors will be to ensure compatibility from one Call Agent to another. Call Agent interoperability is one of the components that could keep service providers from using large-scale, packet-based voice networks.

The service or application plane is where the innovation in the network will happen. One major issue affecting the service plane is its reliance upon soft-switch vendors to open APIs that are useful enough to develop services. For this reason, you will see many application vendors attempting to develop Call Agent technology until APIs into the top Call Agent vendors are fully open and service-friendly.

The service plane is where thousands of ISVs will converge to develop new and revenue-enhancing applications. This is comparable to the client/server revolution in which Microsoft removed the barriers of having to code video drivers, and so on, and enabled ISVs to concentrate on applications. This same revolution is happening in the PSTN today and will change the way services and telephony/multimedia networks are designed, built, and deployed.

Summary

Voice in the PSTN is a fairly complex weave of different technologies that have been evolving since 1876. The PSTN as you know it today is on the verge of a revolution.

The technology required to enable true multimedia conversations on a daily basis is readily available. Such a feat does not require a computer as you know it today. Rather, the telephone/communications infrastructure is moving to a new model and will soon be able to carry these multimedia conversations.

The remaining piece of the puzzle is the bandwidth necessary to complete these multimedia conversations. This is being accomplished in the bandwidth wars currently being fought by the DSL and cable providers. In the end, consumers will be the ultimate winners, in that they will have access to technology that will eliminate distance barriers and communication barriers, and will truly revolutionize the way things are currently done.