Chapter 1

Principles of Converged Networks

The concepts of network convergence — using one network to transmit both voice and data information — are not new. You can trace the evolution of this technology back several decades to the birth of the Integrated Services Digital Network (ISDN). This work, which was standardized by the International Telecommunication Union — Telecommunications Standards Sector (ITU-T), formerly known as the Consultative Committee for International Telephony and Telegraphy (CCITT), was the subject of much discussion and interest in the 1980s. This book begins with a brief look at the history of the topic.

1.1 The Promise of Network Convergence

In the 1980s, the term *convergence* was rarely used. Instead, telephone providers and subscribers used the term *integrated* to describe their vision of using a single telephone line to their home or business as a multifunction line for a variety of applications. These included voice communications (one or possibly two conversations over that single line), data applications for remote host access (stay-at-home shopping, remote electric meter reading, and so on), and even video applications (*à la* the famed Picturephone developed by Bell Telephone Laboratories in the 1950s). At that time, ISDN proponents imagined that these new technologies would displace the currently deployed analog Public Switched Telephone Network (PSTN) within a few years.

Unfortunately, the implementation of ISDN was more of a challenge than the architects of the technology envisioned. When the telephone companies began to deploy ISDN, it was discovered that a large number of the local loops — the physical circuits from the telephone company central office to the end-users' premises (home or small business) — would not support this high-speed data transmission. Some of the loops were too long, and some included systems that optimized the network for voice transmission and would not support the transmission of high-speed data.

As a result, ISDN deployments were focused on the requirements for new construction. For example, if a new housing development required new telephone service, then making that new service compatible with ISDN technology made sense. On the other hand, retrofitting existing outside plant systems to support ISDN was much more expensive and was often not undertaken. As a result, large parts of the United States still do not have ISDN service available. We could say that the promise of converged networks based on ISDN technologies met with limited success. Research from the North American ISDN Users' Forum (NIUF) is a good resource for those interested in ISDN technologies [1-1].

But before dismissing converged voice and data networks as just one more vendor architecture or carrier offering, we need to look at another significant trend, this time coming from the data side of telecommunications — the emerging ubiquity of the Internet Protocol, or IP. This protocol was originally designed in support of U.S. military and higher education data networks and was migrated into general business usage in the 1990s. The predecessor network to the current Internet was named the ARPANET, an acronym for the Advanced Research Projects Agency Network, which was funded in part by the U.S. government. The ARPA architecture divided the computer communication functions into four different layers as shown in Figure 1-1: the Network Interface or Local Network Layer, the Internet Layer, the Host-to-Host Layer, and the Process/Application Layer.

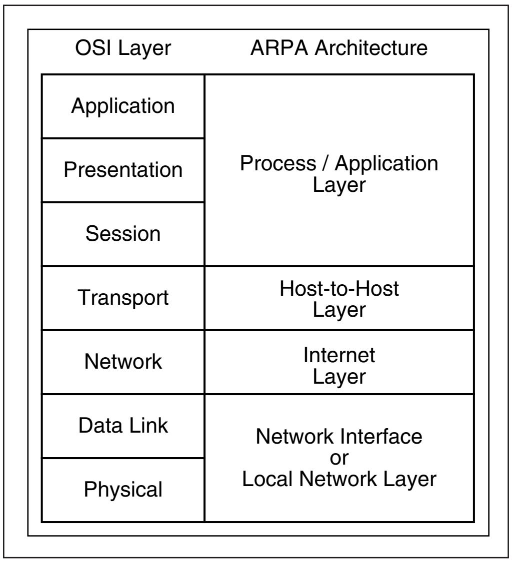


Figure 1-1: Comparing OSI and ARPA Models

Figure 1-1 also compares the ARPA architecture, originally developed in 1969, with the more familiar Open Systems Interconnection (OSI) Reference Model, originally published by the International Organization for Standardization (ISO) in 1978 [1-2].

The ARPA Network Interface or Local Network Layer corresponds with the OSI Physical and Data Link Layers and comprises the hardware elements of the networking infrastructure. The ARPA Internet Layer corresponds with the OSI Network Layer and is responsible for routing and switching the information (typically divided into packets) through that networking infrastructure; it must deal with issues such as packet delivery, routing tables, and addressing schemes. The Internet Protocol (IP) resides at the ARPA Internet Layer. The ARPA Host-to-Host Layer corresponds with the OSI Transport Layer and is responsible for the reliable end-to-end delivery of those packets. The Transmission Control Protocol (TCP) resides at the ARPA Host-to-Host Layer. Finally, the ARPA Process/Application Layer deals with the functions in support of the end-user's application, such as the logical connection (OSI Session Layer), data formats (OSI Presentation Layer), and application-specific support (OSI Application Layer).

At the present time, it would be very difficult to find any type of computing platform, from the smallest cell phone or personal digital assistant (PDA) to the largest mainframe computer, that does not support IP. In most cases, IP is used in conjunction with other protocols, such as the Transmission Control Protocol, or TCP. The term TCP/IP actually refers to a suite of protocols, which in most cases are incorporated into the host's operating system. A companion text, *Troubleshooting TCP/IP* [1-3], provides details on this widespread support for TCP/IP.

Two trends have emerged thus far. If you look at the voice networking side first, you find the idea of some type of combined voice/data network is not new. Unfortunately, the early implementations, such as ISDN, were somewhat problematic; as a result, most of the voice networking infrastructure within the PSTN of today is much like it was several decades ago (allowing for improvements in network technologies and capacities, such as when the PSTN became an all-digital transmission medium around 1990). Consultant James Cavanagh describes this evolution of the telephony system as three distinct markets: the First Wave, that of traditional, highly reliable yet legacy telephone systems; the Second Wave, which introduces Internet Telephony services; and the Third Wave, with its converged voice/data/video systems [1-4].

If you look at the data networking side next, you see that IP has become the common denominator infrastructure of choice and is widely implemented around the world. Thus, addressing the needs for combined voice/data networks, with technologies that are now more implementable than those of decades past, has definite appeal. A number of parties are interested in the success of such developments, including end users, network managers, equipment vendors, and networking carriers. Internetworking between the existing legacy networks and the IP networks, carrying voice, data, video, or other media, will require some careful planning, as Reference [1-5] describes.

This study of Voice over IP systems (VoIP) begins with a look at the principles of network convergence — the concepts that allow voice, data, fax, and video signals to share a common networking infrastructure. I start this journey with a look at some of the principles — and caveats — that are driving this area of technology, as discussed in References [1-6] through [1-12].

1.2 Connectionless vs. Connection-Oriented Network Architectures

In general, communications networks can provide one of two different types of network services: connectionless (CON) network service or connection-oriented (CO) network service. More specifically, an individual layer within a network architecture can be defined by the service it provides to the adjacent layer above: either CO or CON service. Taken as a whole, the communication architecture may incorporate some combination of these services in support of a particular application.

The typical model for a connection-oriented network service is the Public Switched Telephone Network (PSTN). When the end user takes the telephone off-hook, they notify the network that service is requested. The network then returns dial tone, and the end user dials the destination number. When the destination party answers, the end-to-end connection is confirmed through the various switching offices along that path. A conversation (or more typically a voice mail message) can then ensue, and when the conversation is completed, the parties say goodbye and hang up. The network then disconnects the call, terminates the billing process, and makes the network resources available for another conversation.

For the duration of the connection (or conversation), the communication path may be modeled like a very long pipeline: information is inserted into one end of the pipeline and taken out at the other end of the pipeline (Figure 1-2a). During the time the connection is active, certain statements can be made about that pipeline. For example, the data will arrive in the order in which it was sent (*sequentially*). The data will follow the same path (either a physical or logical path, depending on the network architecture in use) from the source to the destination. Finally, the relative delay of the data will be constant along that path. That is, if there is a 20 millisecond delay on the first word that is transmitted through the telephone network, there should be the same relative delay for the second word, the third word, and so on. Since these characteristics, such as sequentiality, delay, and so on, positively impact the quality and reliability of the transmission, a connection-oriented network is often referred to as a *reliable* network. The Transmission Control Protocol (TCP) is an example of a connection-oriented protocol.

In contrast, a connectionless network could be modeled by the postal system. A full source and destination address is attached to the packet (or envelope), and then that information is dropped into the network (or post office box). Each of these packets is routed independently through the network, and through the miracles of packet delivery protocols, delivery to the ultimate destination occurs—we hope, as shown in Figure 1-2b. But, like the postal system, connectionless network delivery is on a "best efforts" basis. This means they will do all they can to get your information through the network, but there are no guarantees of packet sequentiality, delay, or, for that matter, delivery at all. (At least, no guarantees for the price of a first class stamp. If you want more reliable package delivery, you can call FedEx, but it will cost you considerably more. Similarly, if a connectionless network does not meet your needs, you can always go with a CO system, but the overhead will invariably be higher.) The Internet Protocol (IP) and the User Datagram Protocol (UDP) are examples of connectionless protocols.

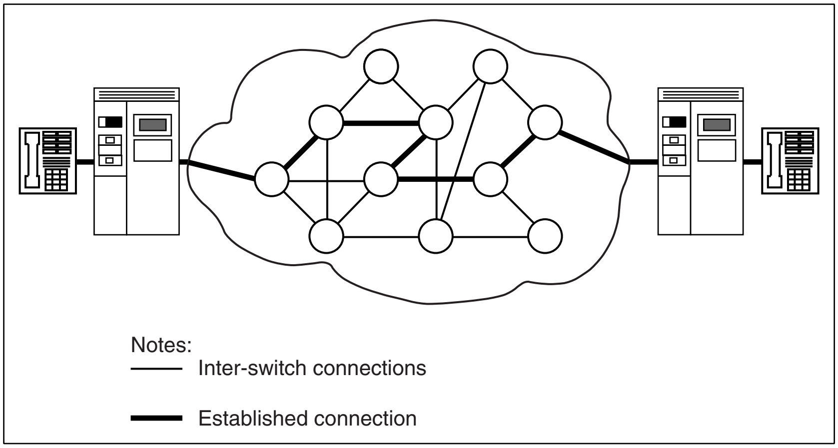


Figure 1-2a: Connection-Oriented Network

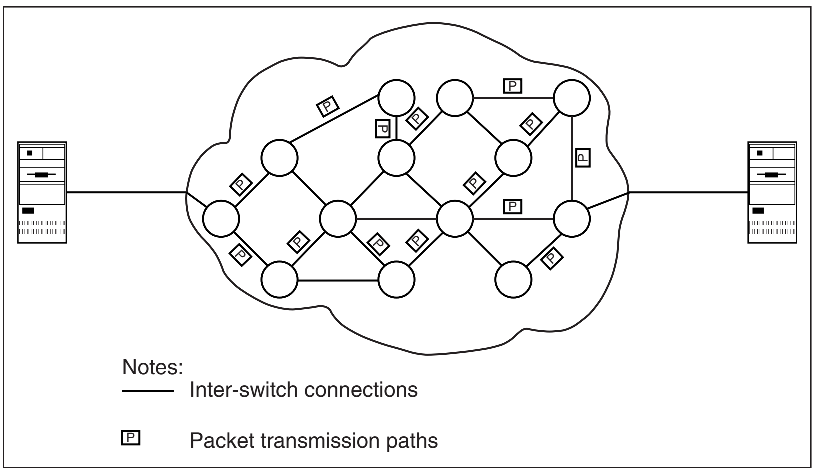


Figure 1-2b: Connectionless Network

1.3 Voice and Data Network Characteristics

Most enterprise networks are really a "network of networks," with distinct infrastructures that address specific requirements. For example, there may be a separate network for circuit-switched voice and fax; a private tie-line network for intracompany voice and fax; a centralized processing data network, such as one based on IBM's System Network Architecture (SNA); a distributed processing data network supporting client/server applications; a private internet or intranet for communication between employees; a public internet or extranet for communication with customers and/or suppliers; and systems and networks supporting other remote access configurations.

In general, voice networking infrastructures are connection-oriented, while data networking infrastructures are connectionless. But in order to determine which network type will be most prevalent in the years to come, consider some specifics of each technology.

1.3.1 Voice Network Characteristics

Traditional voice networks, such as the PSTN, are connection-oriented. To initiate a call, the end user takes their telephone off-hook, which signals the telephone company Central Office (C.O.) that service is requested (Figure 1-3). The end user then enters the destination telephone number via the numeric keypad (or rotary dial, if you are into antiques!). The destination telephone number becomes input to the signaling system among the various C.O. switches to set up the call along the path from the source to the destination. The last C.O. in the chain signals the destination user's telephone that an incoming call is in process by sending ringing current through the line, thus creating an audible signal. When the end user (or that person's answering machine) takes the phone off-hook, the circuit is completed and communication can proceed.

One of the key elements that makes the PSTN so effective is its ubiquity, or universal service—the premise that states that basic telephone service should be an affordable commodity to anyone who wants it. As a result, reaching (almost) any end user via the PSTN should be possible.

The current PSTN creates circuits that provide a bandwidth of 64 Kbps, an element known as a Digital Signal Level 0 (or DS0) channel. The characteristics of DS0 channels are explored in greater detail in subsequent chapters. For now, suffice it to say that the 64 Kbps of bandwidth that is provided with a DS0 (which dates back several decades) is much more than is required by current voice and fax technologies. For example, a fax transmission uses only 9.6 or 14.4 Kbps of bandwidth, while the balance goes unused. Chapter 3 discusses in more detail how this "leftover" bandwidth can be used for other conversations with voice compression technologies, thus providing some of the economies of scale that help justify the investment in VoIP and FoIP equipment.

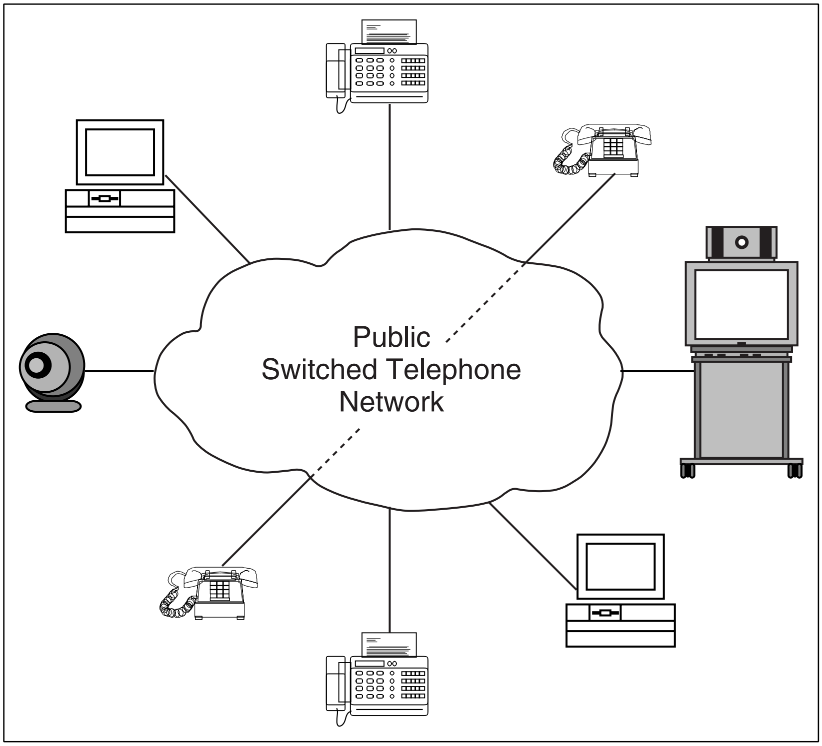


Figure 1-3: Public Switched Telephone Network

In some cases, switched network connections, which provide 64 Kbps per channel, are not adequate for the amount of information that needs to traverse the connection. For example, two PBXs may have a required number of simultaneous connections between two locations. For those applications, a private voice network can be configured, with trunk circuits provisioned according to the bandwidth requirements (Figure 1-4). As an example, if a dozen or so simultaneous conversations were expected between two locations, a Digital Signal Level 1 (or DS1) circuit would likely be employed, which provides 24 of the DS0 (64 Kbps) channels. As one might expect, the bandwidth of the trunk circuit is primarily limited by the depth of your checkbook — the carrier is quite content to provide more bandwidth as long as you are willing to pay for it. For example, the next step in the digital multiplexing hierarchy is known as DS3, which operates at 44.736 Mbps and provides the equivalent of 672 DS0 channels between two locations. Not too many organizations need this amount of bandwidth just for voice, but when other media such as data, fax, and video are considered, that amount of bandwidth is much easier to justify.

For the duration of a particular call, however, resources along the path of the circuit have been reserved on behalf of the communicating parties; as a result, the parties pay for the network services for the time that these resources have been reserved. The reservation of resources provides some predictable characteristics of that connection, including a constant delay for that unique path and sequential delivery of the information (the first word that I speak is the first word that you hear, and so on).

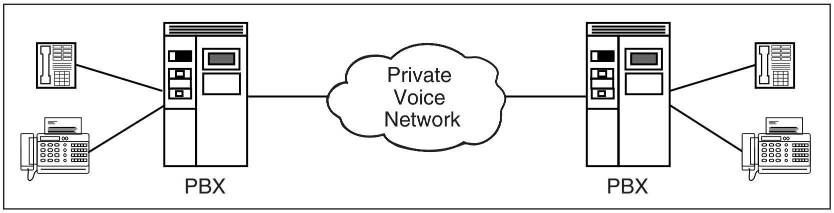


Figure 1–4: Private Voice Network

But if we were to summarize the characteristics of the telephone network, the term *reliable* would have to be used. Since the PSTN is considered a national resource, regulatory bodies such as the Federal Communications Commission (FCC) and state Public Utility Commissions (PUCs) monitor a number of PSTN operational characteristics. These include the number and duration of service outages, the time it takes to repair and restore service in the event of an outage, prices the carriers charge for their services, and so on. The phrase “five nines of reliability” comes from an industry standard for central office switches, which are designed for two hours of downtime in a period of 40 years of service. If all goes according to design, the network downtime would be calculated as follows:

2 hours of downtime/40 years * 365 days/year * 24 hours/day = 2/350,400 or 0.00000571

If this number is converted to availability (or uptime) by subtracting it from 1, you have the following:

Network availability = 1 - 0.00000571 = 0.99999429

Converting this to a percentage yields

99.999429 percent availability (or reliability)

Hence, the phrase “five nines of reliability.”

As we explore throughout this book, the ultrareliable performance of the PSTN can be both a blessing and a curse. It is a blessing in that it rarely fails. (For example, when was the last time you went to a pay phone to make a call, and the pay phone was out of order? In contrast, when was the last time your PC locked up? It has been said that the good thing about conventional telephone service is that you don’t have to reboot the end-user equipment!) It is a curse in that the end users know how reliable the telephone network is and expect that same level of service from a VoIP network.

1.3.2 Data Network Characteristics

In contrast to voice networks, which are generally connection-oriented, data network infrastructures are typically connectionless. Note that a distinction needs to be made between the networking infrastructure and the applications that run over that infrastructure. The applications may require a connection orientation, which would typically be handled by the ARPA Host-to-Host Layer (or OSI Transport Layer). For example, the File Transfer Protocol requires some degree of reliability and synchronization between the sending and receiving file transfer processes. Those functions are handled by TCP, which runs at the ARPA Host-to-Host Layer. IP, which is a connectionless protocol and runs at the (next lower) ARPA Internet Layer, would be considered the defining infrastructure protocol.

The origins of data networking are frequently traced to the development of packet switching and the X.25 protocol in the late 1960s. Packet switching technologies were developed, in part, in response to U.S. government requirements for secure and survivable military communication. The premise was to divide a large message into a number of smaller elements (called *packets*) and to then send these packets via different routes. The diverse routing made it much harder for an adversary to eavesdrop on the message when that message was broken up into smaller pieces and transmitted via different paths (thus yielding some degree of network security). In addition, routing around network failures was also possible, which made the network more robust or survivable. Other benefits were derived from packet switching, such as load balancing between different routes.

Early data networks relied heavily on leased line and packet switched (using X.25) connections, migrating to Internet connections as that transmission medium became more prevalent (Figure 1-5). From the host-to-host applications that were prevalent in the 1960s and 1970s, local area networks (LANs) became the data network of choice in the 1980s. Private data networks that comprised both the local network infrastructure (such as Ethernet LANs) and the wide area connections became the next step in the evolutionary process (Figure 1-6).

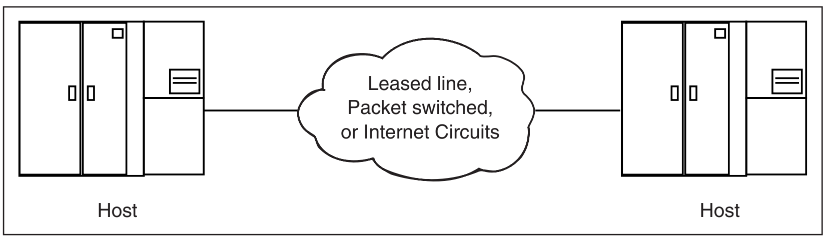


Figure 1-5: Public Data Network

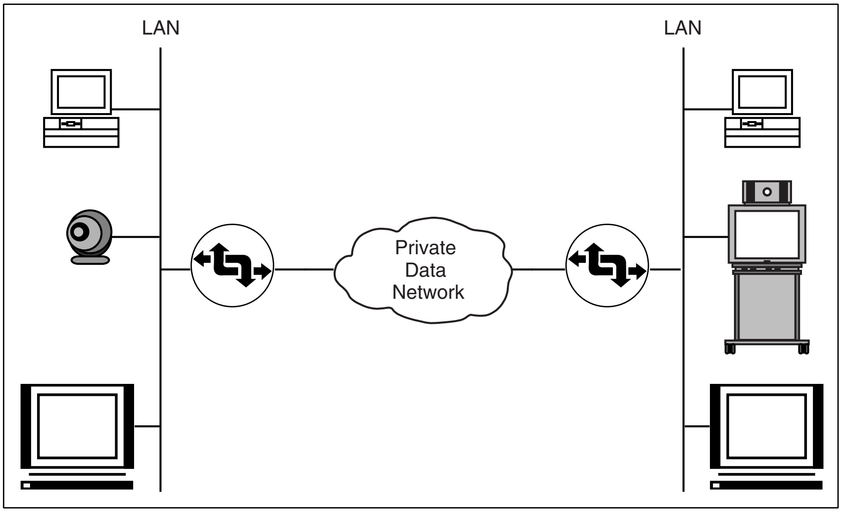


Figure 1-6: Private Data Network

In contrast to the ultrareliable, ubiquitous PSTN, data networks offer what is called "best-efforts" service. In other words, if your transmission on the Ethernet collides with mine, both are destroyed, and the retransmission mechanism built into Ethernet must be employed for subsequent transmission attempts. But that's the way it goes — the Ethernet undertook its "best efforts" to get the transmission through. At best, both transmissions will be delayed. Packet switching networks are another example of best-efforts service, because packets that arrive in error or too late may be discarded, requiring the upper layer protocol to request a retransmission. These multiple attempts yield overall network reliability that is quite a bit less than the 99.999 percent reliability provided on the PSTN. (If you doubt this, ask your end users if they can remember the last time a pay phone failed. Then ask them if they can remember the last time their local network failed. Compare the results, and I think you get the point.) But when the transmission is successful, you have access to a huge amount of network bandwidth — perhaps megabits per second of bandwidth — and are not limited to the 64 Kbps channels that are typically provisioned in voice networks.

Data networks are typically priced based on consumption. If you migrate from 10 Mbps Ethernet to Fast Ethernet to Gigabit Ethernet, you pay more per port. Similarly, if your WAN connections migrate from a DS1 circuit to a DS3, you also pay more. In contrast, PSTN pricing tends to be relatively flat, because the bandwidth consumed is a consistent 64 Kbps. For example, your PSTN rate may be $0.05 per minute any time of the day, to any destination in the United States.

Finally, there is a big difference in the speed at which innovation is required within data networks. These technologies move very quickly. The sales, marketing, engineering, customer support, and other departments within data networking organizations must be prepared to move quickly as well, as the end users of this service demand higher and higher data transmission rates, amounts of memory and information storage, and so on. Contrast this with the innovation that is required in the PSTN. The process of making a telephone call today is about the same as making one several decades ago. The PSTN infrastructure, such as the internal signaling systems, has markedly improved. However, since many of these improvements are not readily visible to the end user, the service presented to those end users may appear to be quiescent.

This brings us back to the subject of network convergence, and the implementation of an integrated voice/data network, such as that shown in Figure 1-7. Note that elements from both public and private voice and data networks are present. Our objective, however, is to take the most favorable characteristics of each network type and to design the new system to optimally handle both voice and data transport with equal facility. Our network infrastructure of choice is based on IP.

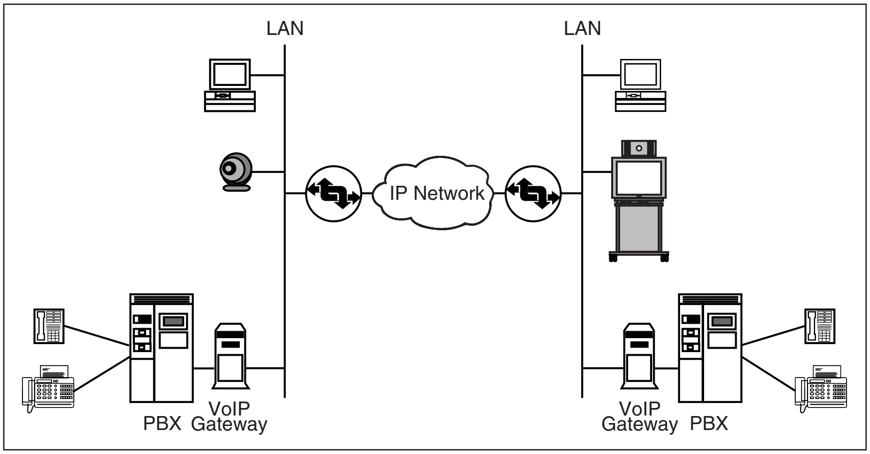


Figure 1-7: Integrated Voice/Data Network

1.4 Voice and Data Network Growth Factors

This section presents market growth and sizing information that relates to the Internet and other IP-based networks. It also contrasts traffic characteristics of voice and data networks. As with most research of this nature, the observations of the general *trends* become much more important than the finite *details* of the projections themselves, as it is unlikely that any of us can claim a high degree of success in predicting the future. Keep that premise in mind as you study this section, and look for the information that can guide you toward an informed decision regarding network convergence plans for your infrastructure.

In most organizations, the *growth* of voice traffic has been relatively flat in the past few years, while the *growth* of data traffic has increased much more aggressively, as shown in Figure 1-8. Note that at this point in our discussion we are considering growth patterns, not actual traffic volumes. Also notice that the exact point of crossover on the time axis has been omitted, as this would be a network-specific characteristic that varies from enterprise to enterprise. For example, a high-technology research center with an ATM switched infrastructure and fiber optic trunk connections may have seen its growth in data traffic exceed that of voice traffic some time ago. In contrast, a call center or service operations center may continue to have more voice than data and may not have an equivalent amount of voice and data traffic growth (the crossover point) for some time to come. And when you consider that most voice traffic today is still 64 Kbps traffic — not compressed, as the technologies of this text discuss — then the amount of traffic (measured in bits per second) that leaves your office may be greater in the voice case (today) than in the data case. In any event, however, consider that your data traffic growth should exceed that of your voice network at some time in the future, hence your interest in a converged network infrastructure.

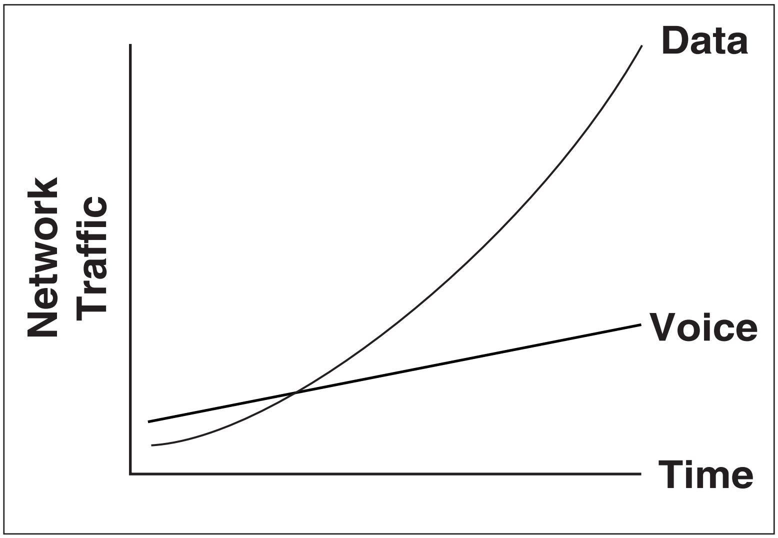


Figure 1-8: Typical Voice and Data Network Growth Patterns

To evaluate this premise, consider the following generalizations. The aggregate voice network (considering the PSTN, private networks, and so on) supports one primary application (voice) and a few data applications (such as modem traffic) that consume relatively small amounts of bandwidth (typically 64 Kbps per channel or less). The aggregate data network (including the Internet, private internetworks, virtual private networks, and application-specific networks, such as SNA backbones) supports many data applications. Much of this is heavily influenced by the large growth in the number of users attached to the Internet. The most prominent of this Internet traffic is World Wide Web information, which is clearly growing at exponential rates, fueled by electronic commerce and other interactive applications. (For an interesting look at how fast the number of Internet-connected hosts is growing on a daily basis, consult Reference [1-13].) It would not be an exaggeration to depict the generalized growth in voice traffic as being relatively flat (or linear), while the growth in data traffic has some positive slope (and possibly an exponential slope), as depicted in Figure 1-8.

The Yankee Group (Boston, Massachusetts) estimated the annual growth rate of traffic over a three-year period, and reached some of the same conclusions [1-14].

Traditional (PSTN) voice traffic is projected to decrease in North America and to increase slightly in Europe (Figure 1-9a). Trends for traditional data traffic are similar, but with larger variations (Figure 1-9b). The Internet/IP traffic projections tell a different story, however, as their increase is more than double that of the other two media types (Figure 1-9c). Therefore, these results in Figures 1-9a, b, and c quantify the intuitive hypothesis illustrated in Figure 1-8, supporting the idea that traffic will migrate from traditional networks, such as the PSTN, to IP and converged network infrastructures.

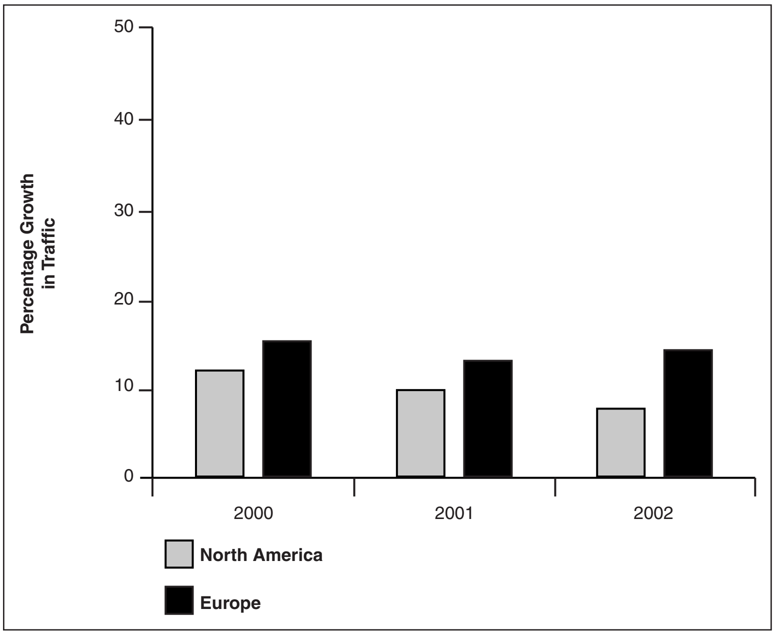


Figure 1-9a: Estimated Annual Growth Rate of Traditional Voice Traffic

(Copyright 2001, The Yankee Group)

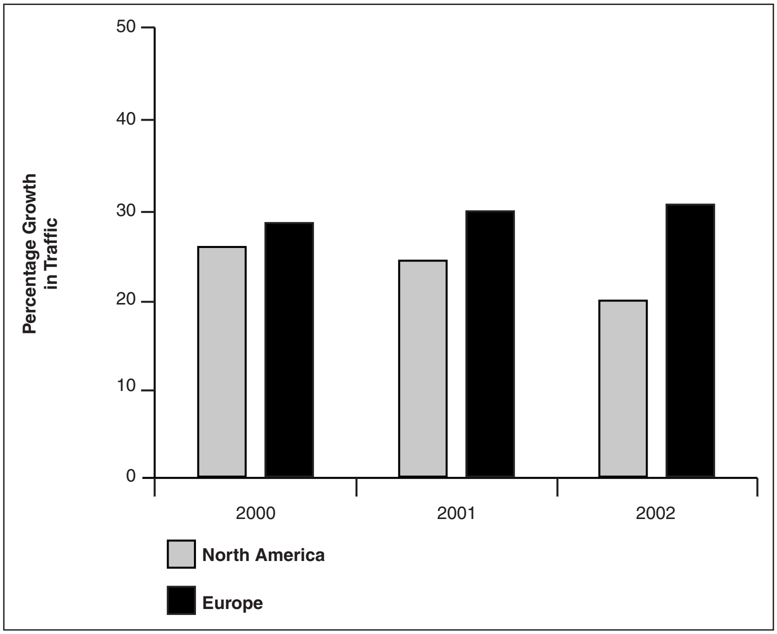


Figure 1-9b: Estimated Annual Growth Rate of Traditional Data Traffic

(Copyright 2001, The Yankee Group)

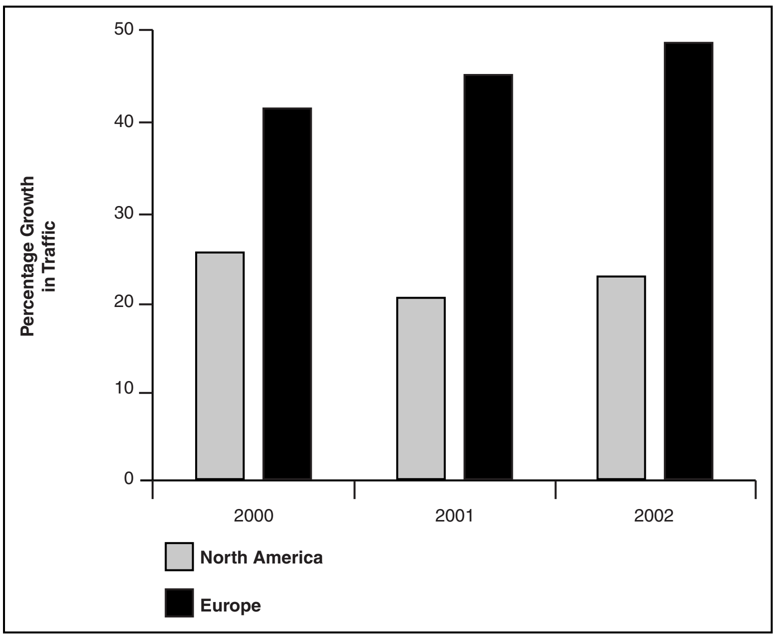


Figure 1-9c: Estimated Annual Growth Rate of Internet/IP Traffic

(Copyright 2001, The Yankee Group)

An interesting paradigm shift is occurring. In the last several decades, data has been treated as a special application to be sent over voice circuits, with dial-up modem traffic the most prevalent example. In today's environment, with packet-oriented data traffic being the area of highest growth, treating the voice traffic as a special application of data transport garners some appeal.

This can also be illustrated when the mix of global network traffic is segmented by both public (PSTN) and packet/private networks (Figure 1-10). As discussed earlier in the chapter, the PSTN voice traffic has relatively flat growth (the lower portion of Figure 1-10). When packet and private voice traffic is considered, network traffic is only slightly higher (the next highest, or second portion of Figure 1-10). PSTN data traffic has a higher rate of growth than either of the voice cases (the next highest, or third portion of Figure 1-10), while the packet and private data traffic has the highest growth of all (the top portion of Figure 1-10), once again supporting the hypothesis that traffic is migrating from the PSTN to packet-based networks that include a mix of both voice and data traffic.

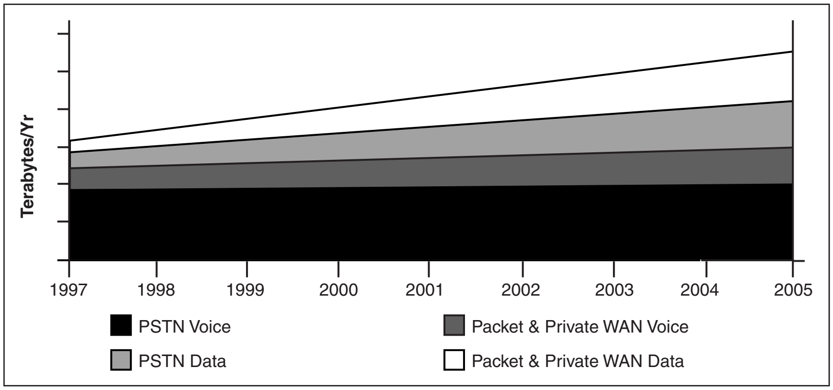


Figure 1-10: Global Network Traffic

(Courtesy of Probe Research, Inc.)

New public network infrastructures are being designed and built to handle these consolidated networks, by companies such as Quest Communications International, Inc., and Level 3 Communications, Inc. In building these new infrastructures, these firms are leveraging some of the significant increases in technology that have occurred in the last few years, such as high-capacity fiber optic backbones and routers that can handle aggregate throughputs in the gigabit per second range. These increases in technology provide one key ingredient necessary for the success of converged networks—favorable economies of scale that promise to reduce the overall costs of the enterprise infrastructure.

But determining the sources of this growth is essential for understanding the rate at which this growth will occur. Probe Research has identified five different market drivers for the Voice over IP (VoIP) and Fax over IP (FoIP) market, along with projected time frames when these market drivers will become significant (Figure 1-11). The Hobbyist phase (1996–1997) allowed two PC users equipped with compatible software and sound cards to send voice over the Internet. The Tariff Arbitrage phase (1998–1999) afforded the end user an opportunity to lower long-distance telephone costs, which became especially important for international connections. Parity with existing public and private voice systems will drive the market in the next phase (which began in 2000). For example, IP telephony gateways must provide interoperable, not proprietary, solutions, and compatibility with existing voice switching systems, such as private branch exchanges (PBXs) and central office (C.O.) switches, has become a higher priority of the vendor community. New applications have driven the market during the following phase (2001–2002), as both end users and network managers realize the benefits of unified messaging, multimedia conferencing, enhanced call centers, and other key applications. The final phase (2003–2005) will provide cost parity between packet-based voice services and the existing circuit-switched voice services.

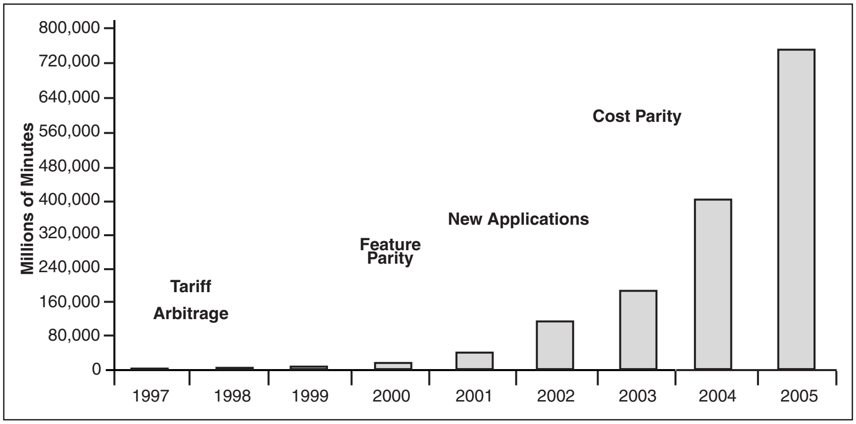


Figure 1-11: Global Voice and Fax over IP Traffic and Market Drivers

(Courtesy of Probe Research, Inc.)

In other words, the costs of the new technologies must present some advantage over the existing, well-established technologies in order to convince network managers of the long-term viability of voice and fax over IP solutions.

During these five phases, the end users of this technology can be segmented into several categories. Probe Research divides the IP telephony market into three general areas: wholesalers of telephony services, consumer/small business users, and enterprise/institutional customers (Figure 1-12). The wholesalers are those who are reselling telephone service via calling cards purchased at discount stores, convenience stores, gas stations, and so on — a market that appears to be growing at a very rapid pace, based on the number of retail displays for these products seen in stores. The calling card purchaser (or end user) is interested in low-cost telephone service, and the more technical issues regarding quality of that service, delays, and so on are of less concern (assuming, of course, that some minimum threshold of acceptable service has been met or exceeded). The consumers and small business users are possibly the most price sensitive, but are also the earliest adopters (note their presence in the 1997–1999 time frame). Enterprise and institutional customers are slower to adopt this technology (the 2000–2001 time frame), but after that have a high rate of growth as the technology is further implemented and tested. One might speculate that concerns over quality make for a more cautious entry into this technology for these organizations.

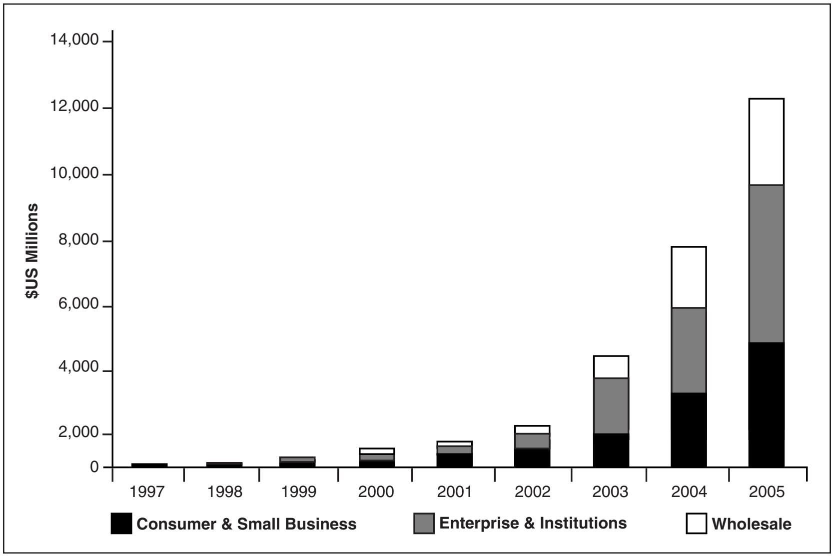


Figure 1-12: IP Voice/Fax Service Markets by Segment

(Courtesy of Probe Research, Inc.)

Financial research firm U.S. Bancorp Piper Jaffray, Inc. (Minneapolis, Minnesota) has looked at the voice and fax over IP market from the perspective of equity investments and has published a very comprehensive guide to these technologies: *IP Telephony: Driving the Open Communications Revolution* [1-15].

For the purposes of their research, Piper Jaffray segments the IP telephony market into five different categories. The Core Enabling Technology Platforms are the hardware and software components that enable IP telephony functions, and include various network interfaces, call processing and digital signal processing capabilities, switching systems, and so on. The Enterprise Solutions category includes voice gateways, client software, and other telephony functions. Applications would enable new end-user features and functions. The Carrier-Class Solutions are systems that are implemented within a carrier's network to include large-scale servers and routers, gatekeepers, billing systems, and so on. End-to-End Services are those organizations that are integrating voice and data equipment sales, installation, and other consulting-related services to other organizations.

Piper Jaffray estimates that, by the year 2003, the aggregate IP telephony market will total $14.7 billion in revenues. This is segmented somewhat differently and is broken down as follows: 6 percent from the core enabling platforms, 27 percent from gateway equipment, 8 percent from applications, and 59 percent from services. Specifics regarding the growth divided by the five segments noted above is shown in Figure 1-13.

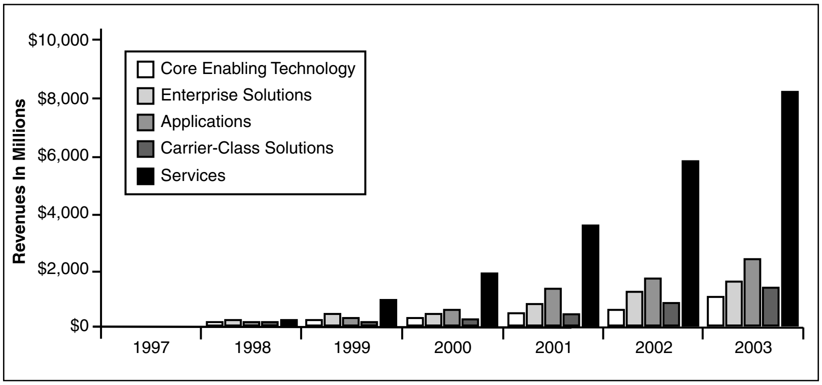
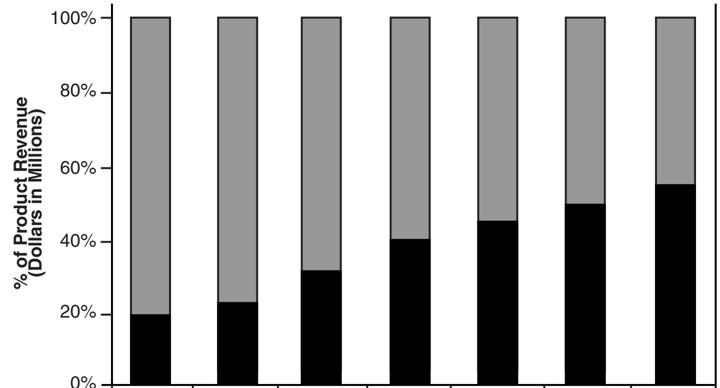


Figure 1-13: Market Growth Forecast 1997–2003 by Segment

(Courtesy of U.S. Bancorp Piper Jaffray, Inc.)

This market can be further segmented by geography, detailing the mix of product revenues that will come from international or domestic (U.S.) sources (Figure 1-14a). Note the shift from the 1997 conditions (20 percent domestic and 80 percent international) to the 2003 projections (55 percent domestic and 45 percent international), as more domestic carriers and firms adopt these technologies.

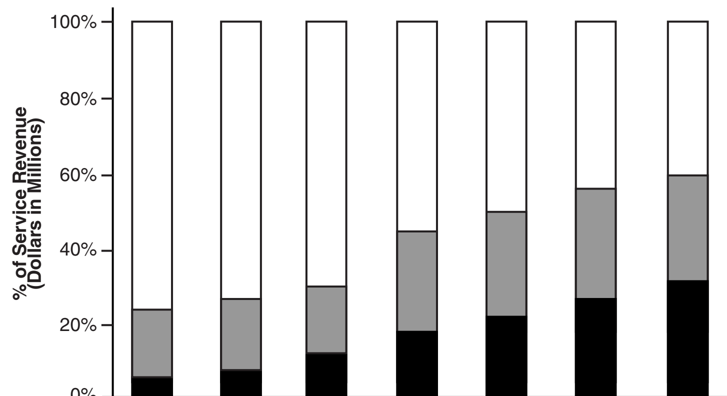
A similar change is noted in Figure 1-14b, which shows the mix of service revenues from the 1997 conditions (5 percent domestic-to-domestic, 20 percent international-to-international, and 75 percent domestic-to-international) to the 2003 projections (30 percent domestic-to-domestic, 28 percent international-to-international, and 42 percent domestic-to-international). As with the previous case, the domestic traffic will increase at the expense of the international traffic.



	1997	1998	1999	2000	2001	2002	2003
International	$48	$231	$462	$785	$1,332	$1,937	$2,723
Domestic	12	69	198	524	1,089	1,937	3,329

Figure 1-14a: Market Growth Forecast 1997–2003 by Geography (Percent of Product Revenue)

(Courtesy of U.S. Bancorp Piper Jaffray, Inc.)



	1997	1998	1999	2000	2001	2002	2003
Domestic-Int'l	$18	$86	$497	$1,017	$1,667	$2,631	$3,664
Int'l-Int'l	5	24	124	509	917	1,608	2,371
Domestic	1	9	89	324	750	1,608	2,586

Figure 1-14b: Market Growth Forecast 1997–2003 by Geography (Percent of Service Revenue)

(Courtesy of U.S. Bancorp Piper Jaffray, Inc.)

The market growth by number of minutes of use for VoIP and FoIP services is also a key indicator of the anticipated popularity of the service (Figure 1-15). IP telephony traffic in 1998 was estimated at 476 million minutes, with 57 percent for VoIP and 43 percent for FoIP. By 2003 the total is projected to be 81.7 billion minutes, with 77 percent from VoIP, 17 percent from FoIP, and 6 percent from video over IP services.

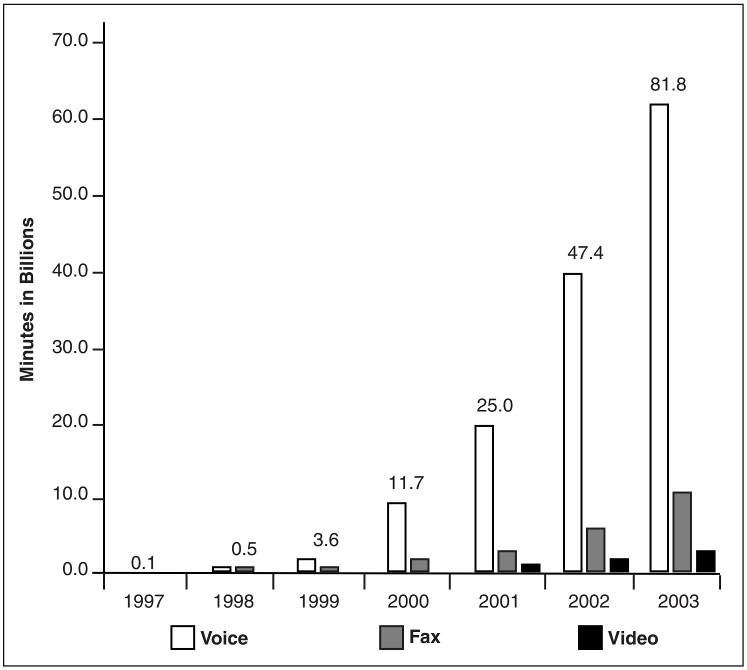


Figure 1-15: Market Growth Forecast 1997–2003 by Minutes Used

(Courtesy of U.S. Bancorp Piper Jaffray, Inc.)

The current market for telecommunication services is very large, dominated by carriers such as AT&T (United States), NTT (Japan), Deutsche Telekom (Germany), and British Telecom (United Kingdom), which primarily offer traditional dial-up and leased line solutions [1-15]. As VoIP and FoIP solutions become more commonplace, it is anticipated that a significant number of the minutes of use (MOUs) will migrate to IP-centric carriers (Figure 1-16). By 2003, it is estimated that this will represent over 81 billion MOUs, and $8.6 billion in revenues (Figure 1-17). Reviewing Figure 1-16, note that, as a percentage of the total PSTN MOUs, those that represent IP telephony services are a relatively small number. However, 6 percent of the total aggregate of worldwide telephony usage still produces some very large revenue numbers! This may be one reason that the more traditional carriers (such as AT&T) are developing IP-centric services to compete with the start-up Internet Telephony Service Providers (ITSPs) such as Level 3 and Qwest. Reviewing

Figure 1-15, note that, in the 2002–2003 time frame, video over IP services will become significant market factors as well.

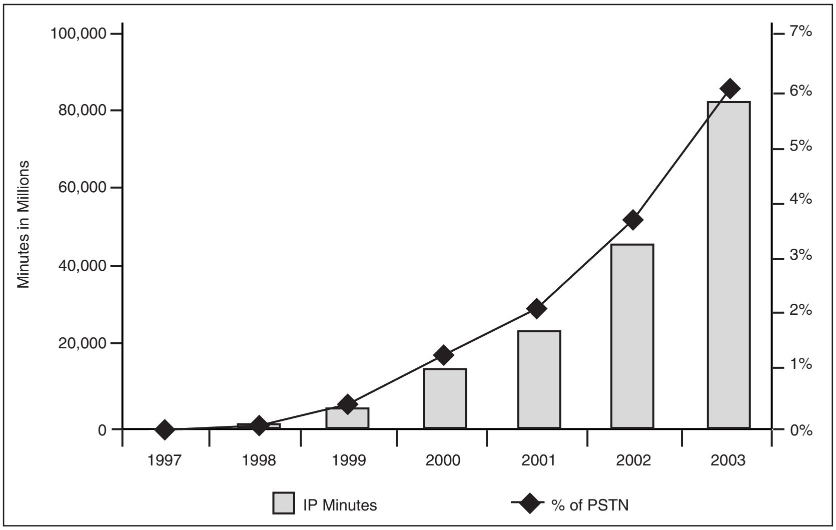


Figure 1-16: Percentage of Long-Distance Minutes Moving to IP Networks

(Courtesy of U.S. Bancorp Piper Jaffray, Inc.)

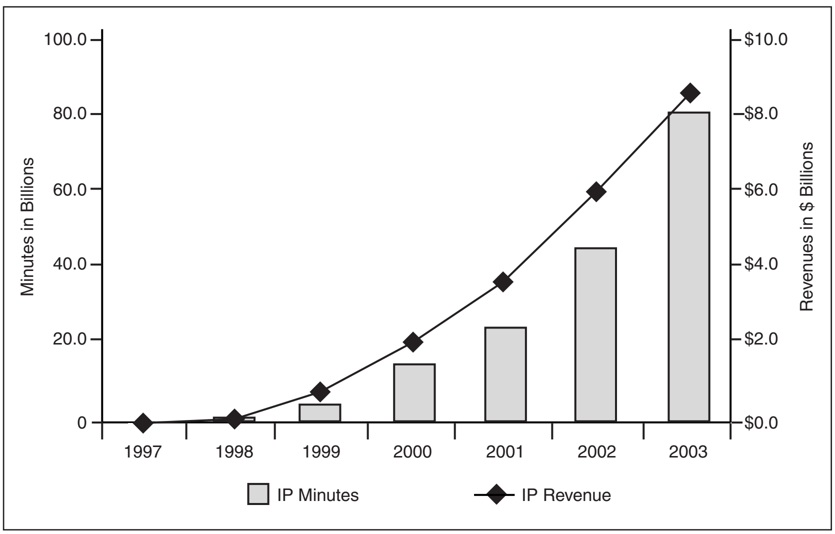


Figure 1-17: Service Provider Revenue and Minute Growth

(Courtesy of U.S. Bancorp Piper Jaffray, Inc.)

1.5 The New Market: Voice over Packet Transport

Thus far, this discussion has focused on Voice over IP technologies, assuming that hardware, software, and communication systems involved in the process are somehow connected to, or derived from, Internet research. However, there is a more general technology sector, known as Voice over Packet, or VoP, which may have a bearing on your converged network infrastructure. The VoP marketplace includes a number of broadband technologies that could be used as an alternative to conventional systems, such as the Public Switched Telephone Network (PSTN) to transport digitally encoded voice information. Examples of these transport options include Asynchronous Transfer Mode (ATM), frame relay, residential broadband (cable television systems), and Packet over SONET (the Synchronous Optical Network), abbreviated PoS.

The significance of the VoP technologies is in the amount of market share that they are projected to take away from conventional PSTN technologies in the next few years. Market researchers Probe Research, Inc., project that VoP will penetrate 14 percent of the traffic on the PSTN by 2006, as shown in Figure 1-18, taken from Reference [1-16]. While the magnitude of this number (14 percent) is not that large, it nevertheless represents a significant amount of revenues to traditional carriers such as AT&T, MCI Worldcom, and Sprint in an era of ever-decreasing profit margins.

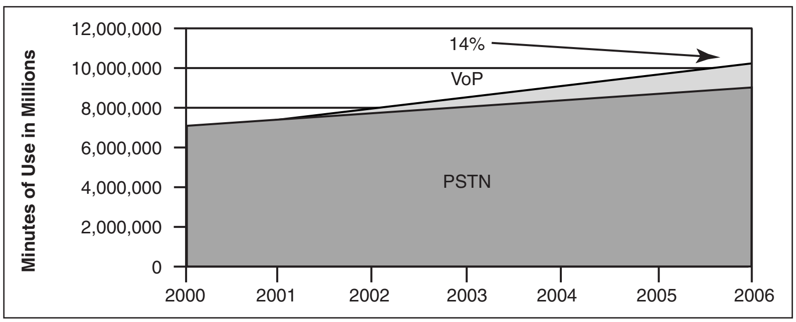
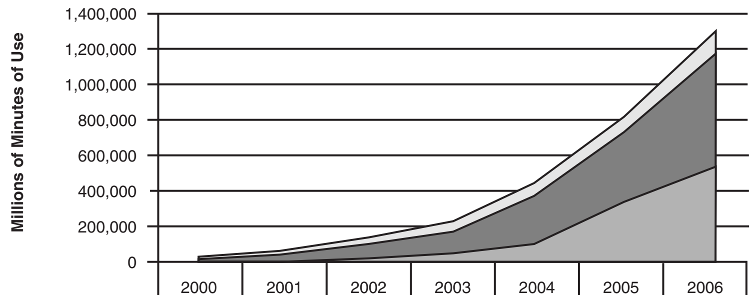


Figure 1-18: Worldwide PSTN and VoP Traffic — 2000–2006

(Courtesy of Probe Research, Inc.)

Voice traffic is measured in *minutes of use*, with voice over packet minutes, as projected by Probe Research, to be any of these call minutes that are packetized for at least a portion of the call along its path from source to destination. Thus, a call that originates on the PSTN and then traverses an IP network would count as a VoP call. Likewise, a call that is packetized at the customer premises, and carried over a public network to its destination, would also be considered a VoP call. This VoP traffic can be segmented into three different markets: international, domestic long distance, and local traffic, as shown in Figure 1-19. Note that international traffic made up the majority of the VoP traffic until 2000, but domestic long distance is projected to dominate the subsequent years. Local traffic has the smallest initial market share, but grows to almost equal the domestic long-distance market in later years. Probe's research concludes that growth in the local VoP segment has been thwarted by the recent industry tumult from both the regulatory and business perspectives, but that growth should occur as those issues are resolved.

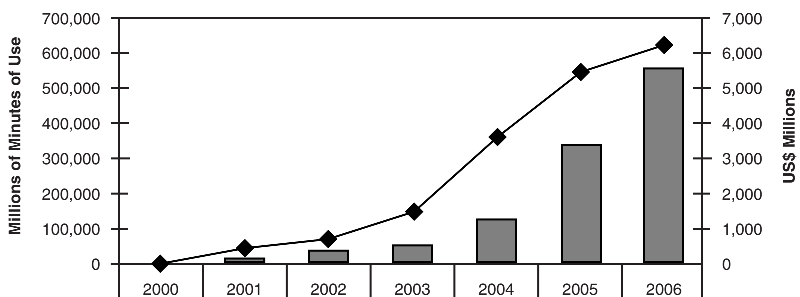


	2000	2001	2002	2003	2004	2005	2006
International	6,223	11,056	23,043	36,576	53,566	76,347	101,274
Domestic Long Distance	4,682	13,038	52,729	137,358	241,371	370,357	603,638
Local	52	4,008	16,273	42,023	123,401	327,113	574,104

Figure 1-19: Worldwide VoP Traffic by Call Segment — 2000–2006

(Courtesy of Probe Research, Inc.)

When considered individually, these three markets project very positive growth numbers. The worldwide local VoP traffic and revenues are projected to reach over 574,000 million minutes of use and almost $6 billion in revenues by 2006 (Figure 1-20).

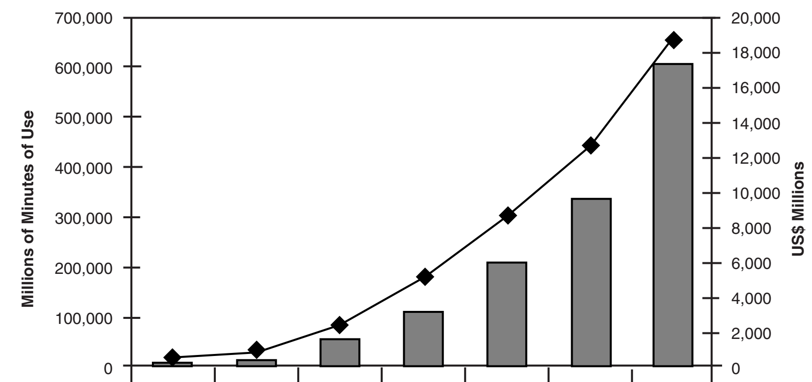


	2000	2001	2002	2003	2004	2005	2006
Traffic	52	4,008	16,273	42,023	123,401	327,113	574,104
Revenues	$2	$184	$656	$1,682	$3,894	$5,452	$6,211

Figure 1-20: Worldwide Local VoP Traffic and Revenues — 2000–2006

(Courtesy of Probe Research, Inc.)

The worldwide domestic long-distance VoP market is projected to reach over 600,000 minutes of use and over $18 billion in revenues by 2006 (Figure 1-21).



	2000	2001	2002	2003	2004	2005	2006
Traffic	4,682	13,038	52,729	137,358	241,371	370,357	603,638
Revenues	$288	$705	$2,550	$5,791	$9,000	$12,223	$18,333

Figure 1-21: Worldwide Domestic Long-Distance VoP Traffic and Revenues — 2000–2006

(Courtesy of Probe Research, Inc.)

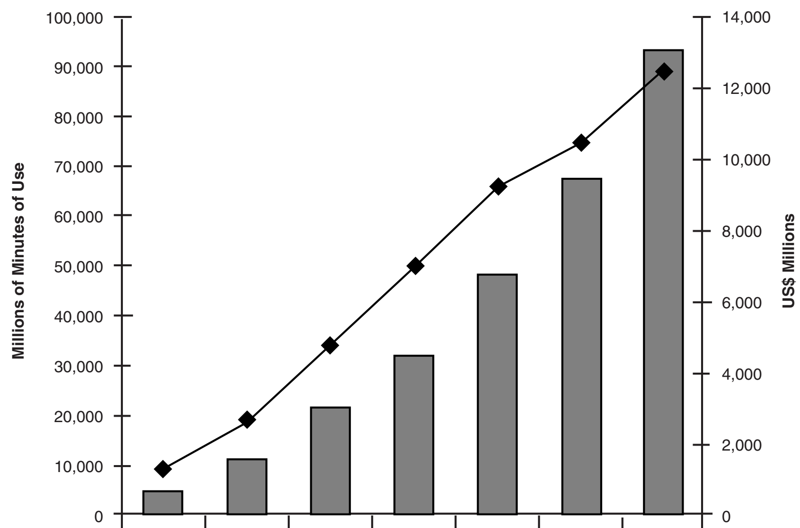
Finally, the worldwide international VoP market is projected to reach over 93,000 minutes of use and over $12 billion in revenues by 2006 (Figure 1-22). In summary, the potential for these converged markets is so large that a number of organizations, from incumbent carriers to equipment vendors, are all vying for a piece of that revenue.

1.6 Benefits of the IP-Centric Network

These are the major hypotheses that have been presented in this chapter thus far:

- The idea of converging voice and data traffic into a single network is not new.
- The Internet Protocol, and IP-related technologies, are well understood and widely implemented (near-ubiquity) within the telecommunications industry.
- For many reasons, data networks, which incorporate packet switching technologies, are generally more efficient than voice networks, which use circuit-switching technologies.

- The growth of traffic from voice networks is relatively flat, in sharp contrast to the growth of traffic from data networks, which is very steep.
- Given the large number of voice communication, data communication, and Internet communication users, plus the large volumes of revenues from these industries, the total market potential for the voice, fax, and video over IP technologies is very large.



	2000	2001	2002	2003	2004	2005	2006
Traffic	6,251	11,000	22,414	33,565	48,635	69,268	93,856
Revenues	$1,679	$2,535	$4,582	$6,907	$9,471	$10,308	$12,412

Figure 1-22: Worldwide International Long-Distance VoP Traffic and Revenues — 2000–2006

(Courtesy of Probe Research, Inc.)

With these hypotheses in mind, consider the following:

Given the current environment, does it still make sense to enhance the capabilities of the voice network to transport data, or, instead, to enhance the capabilities of an IP-centric data network to carry additional applications (voice, fax, and/or video)?

Assuming that such an IP-centric network is deployed, it is important to define the services that it can support. Nortel Networks' document, "Enhancing IP Network Performance" [1-17], divides IP services into three categories: application services, enabling services, and internetworking services. Figure 1-23, which is modeled after the seven-layer reference model, employs an IP-centric system to tie the networking infrastructure at the lower layers to the applications at the higher layers. The three key application services include mission-critical applications, such as human resources, finance, customer support, and so on; voice and fax over IP; and multimedia over IP, such as video conferences, distance learning, and so on. The enabling services enhance the capabilities of IP-based networks, dealing with issues such as traffic management, security, directory services, policy-based networking, and virtual private networks (VPNs) with remote access and extranet support. Internetworking services provide access to the IP-based network from non-IP-centric networks (such as SNA or X.25 WANs). Note that the core of this technology is the Internet Protocol.

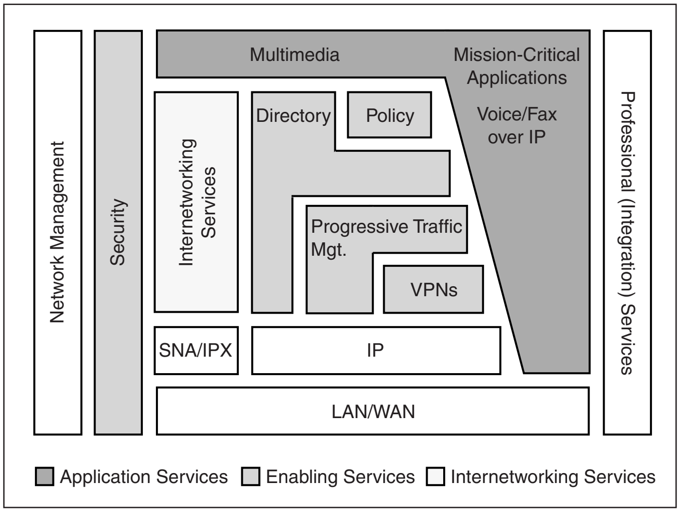


Figure 1-23: IP Services Categories

(Courtesy of Nortel Networks)

The benefits of such a converged network could include:

- A reduction in the overall complexity of the networking infrastructure: fewer protocols, operating systems, and so on.
- Synergies between carrier circuits supporting voice and data, and possible elimination of circuit redundancy.
- Possible reduction in carrier circuit charges.
- Integrated management systems and strategies that can support both the voice and data networks, instead of separate management systems.
- More consistent user interfaces, such as integrated fax and e-mail.
- Access to enhanced applications, such as integrated messaging, voice-enabled Web sites, desktop video conferencing, and others.

The next section considers how the converged network can be deployed.

1.7 Provisioning the Converged Network

Thus, the IP telephony network would include a number of elements—some from traditional voice networks, such as telephones, fax machines, and PBXs; some from traditional data networks, such as terminals, hosts, and servers; and some from new systems, such as gateways and gatekeepers, that are designed as the glue to hold all of these disparate systems together. Figure 1-24 illustrates some of these elements, which could incorporate many different subsystems, including the PSTN, dedicated WAN connections, IP routers, wiring hubs and end stations, and many others.

A call over an IP network could start with a local PSTN connection and the conversion of the analog voice signal to a digital pulse stream at the client end station, PBX, or gateway. That pulse stream will likely undergo some signal processing, including silence suppression and echo cancellation. The gateway may consult with a gatekeeper and/or Domain Name Service (DNS) server to obtain information about the desired destination, such as its transmission capabilities and IP address. The voice signal is then converted into packets, transmitted over the IP-based network to a destination gateway, and the digital pulse stream is reconstructed. As a final step, the digital pulse stream is then converted into an analog signal and delivered to the desired destination.

1.8 Challenges of the Converged Network

Despite the very positive projections for the converged network marketplace, the path from theory and concept to implementation and reality can have a few obstacles and detours along the way. In the case of voice, fax, and video over IP systems, the challenges come in several areas, as discovered by Sage Research, Inc., and illustrated in Figure 1-25 [1-18]. Respondents from large enterprise organizations that currently implement or plan to implement VoIP were asked, "Which of the following are/were challenges to deploying VoIP? Please check all that apply." Almost two-thirds of the respondents indicated that the maturity of the technology (or perceived lack thereof) presented a major challenge to their plans. Other key challenges included convincing executive management of the need for the converged technology, compatibility concerns, and cost issues. (Note that multiple responses were allowed, making the total on the graph exceed 100 percent.)

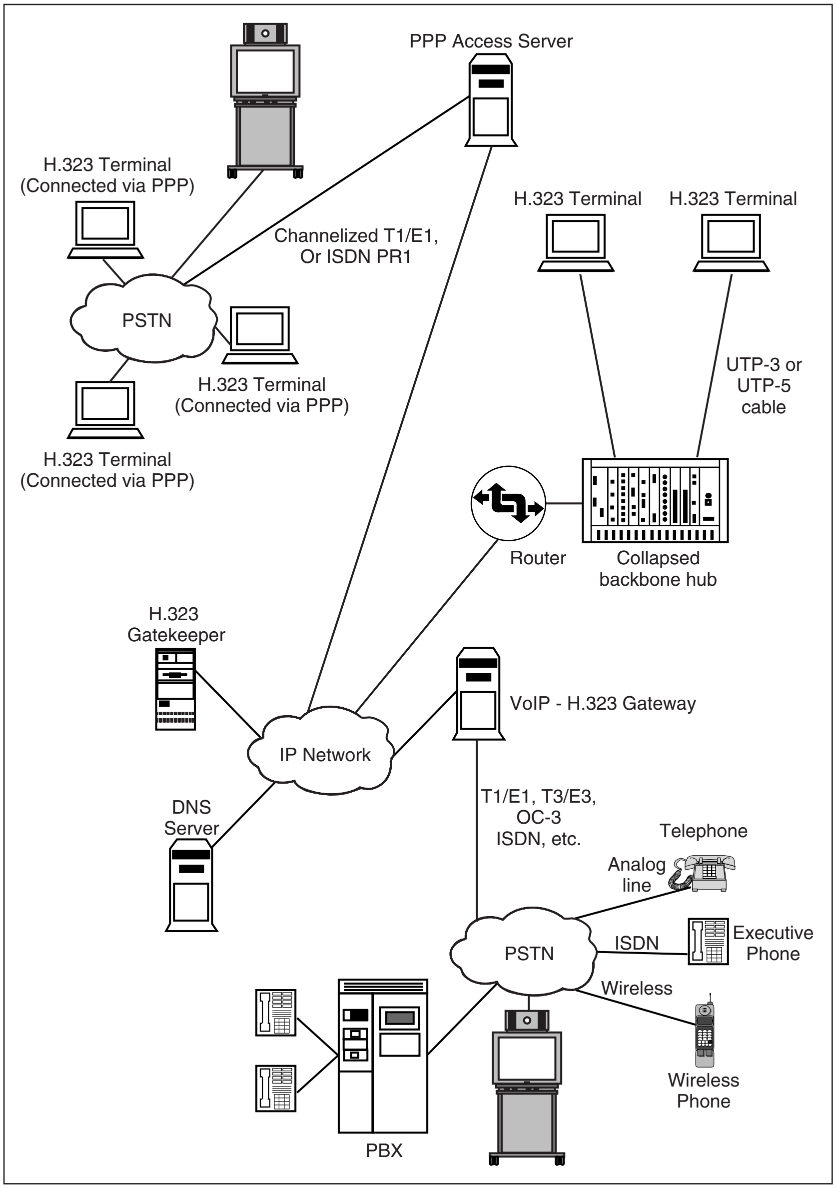


Figure 1-24: Voice over IP Network Elements (Possible Configuration, Not Mandatory)

(Source: Voice over IP Forum)

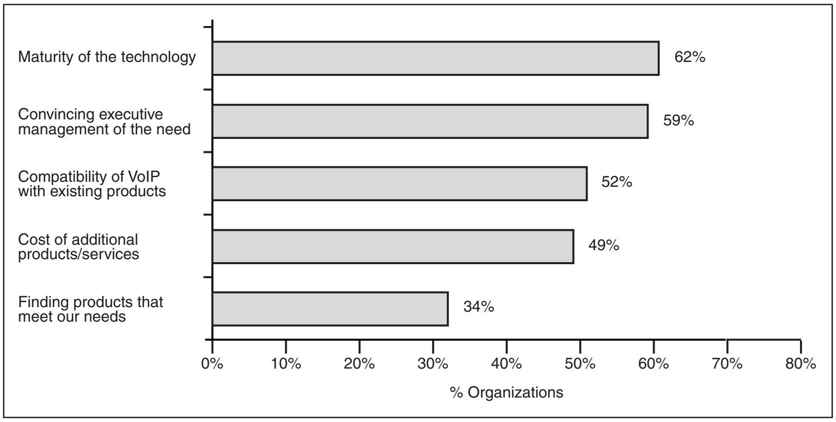


Figure 1-25: VoIP Challenges

(Courtesy of Sage Research, Inc., Natick, Massachusetts)

Those respondents from large organizations that *did not* plan to implement VoIP were asked, "Please indicate the single most important reason your organization is not deploying VoIP at this time." As illustrated in Figure 1-26, slightly more than one-third of the respondents (36 percent) felt that the technology was not mature enough to be deployed, while over one-fourth of the respondents (26 percent) felt that they did not have a need to deploy the technology at this time.

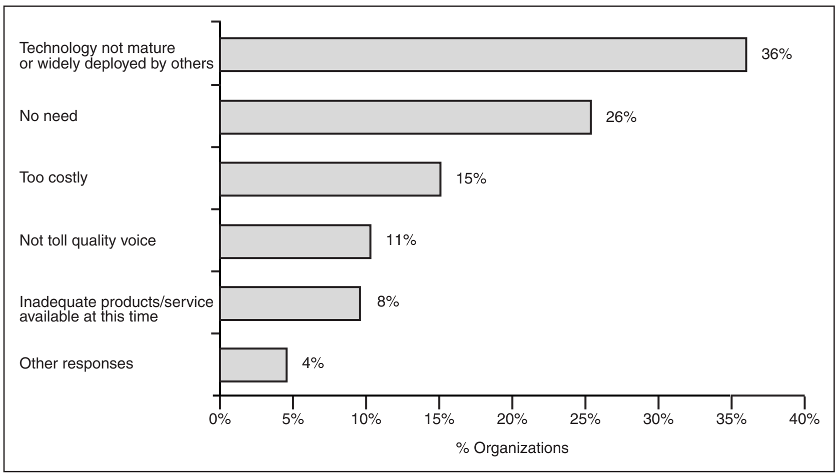


Figure 1-26: Primary Reasons for Not Adopting VoIP

(Courtesy of Sage Research, Inc., Natick, Massachusetts)

Thus, the existing mindsets of the networking constituents, the large number of players in the marketplace, and the technical difficulties afforded by the marriage of voice and data systems all contribute to the challenges of converged networks. The next sections look at these three areas individually.

1.8.1 Voice and Data Networking Mindsets

As was discussed earlier in the chapter, voice and data networks have been designed to address different challenges and, as a result, those who design, configure, and manage these two types of networks approach their respective challenges from different viewpoints. For example, in many organizations, voice telephone service is treated as a utility, much like the electric or water utilities (and possibly taken for granted). As such, it is not always afforded the high-technology moniker that is given to the computing side of the house.

Data networking requirements are often driven by the applications they support, such as a customer database. Thus, the data network is more likely to be associated with the success (or lack of success) of the entire enterprise. (If you doubt this, pick up a networking journal and look at a testimonial advertisement. The typical message is: “We installed XYZ networking products, our downtime decreased, our throughput increased, and we all lived happily ever after. If you buy XYZ products, you will become successful as well.”)

Thus, when one considers implementing an integrated network, which will comprise elements from both the voice and data networks, key issues, such as who will fund, who will manage, and who will be ultimately responsible for the converged network, need to be resolved.

1.8.2 VoIP Market Players

Reviewing Figure 1-13, we note that there are five categories of segments to the IP telephony marketplace: enabling technology platforms, enterprise solutions, applications, carrier-class solutions, and end-to-end systems integrators.

Noting the wide diversity in these five categories, and considering the extensive growth factors that were discussed in Sections 1.4 and 1.5, it is not hard to realize that a number of different organizations have entered, or will soon enter, this marketplace. Examples include:

- Traditional carriers: both Inter-Exchange Carriers (IXCs) and Local Exchange Carriers (LECs)
- Internet service providers (ISPs)
- Internet Telephony Service Providers (ITSPs)
- Fax service bureaus

- PBX equipment manufacturers
- Networking equipment manufacturers
- Application developers

For completeness, we would need to add the standards organizations, such as the Internet Engineering Task Force (IETF), the International Telecommunications Union (ITU), and the Institute of Electrical and Electronics Engineers (IEEE); plus special interest groups, including the Voice over IP Forum and the International Multimedia Teleconferencing Consortium (IMTC), and others, which produce work that crosses many of the above technology segments.

1.8.3 VoIP Implementation Challenges

As we have discussed in this chapter, integrating the significant differences between voice and data networks into a single, cohesive system, and then integrating those systems, can be challenging. To identify some of the topics that are considered in subsequent chapters of this book, as well as to provide some food for thought, consider the following (Figure 1-27):

- Numerous standards exist in support of voice and data networks. How will these standards be applied to a converged network?
- Voice networks are ultrareliable. How will this reliability be translated into the converged network?
- The convergence industry is at its early stages. Will the products and services that you select from different vendors be interoperable?
- Network transmission characteristics, such as latency or delay, are more predictable with voice networks than data networks. How might these differences affect the end-user applications?
- How will the converged network integrate with any existing systems, such as a PBX or voice mail network?
- Does your network infrastructure have enough excess bandwidth to support voice, fax, and/or video applications?
- How will the converged network be managed — from the perspective of voice transmission, from the perspective of data transmission, or as an integrated system?
- Will regulatory bodies, such as the FCC, impact voice transmission over IP-centric networks, such as the global Internet?

- How will the existing levels of Quality of Service (QoS) be maintained? This challenge needs to be addressed from two perspectives: the perspective of the end user who is concerned about the quality of their voice call, and the perspective of the recipient of that call, perceiving whether or not the caller is understandable. This issue is especially important for emergency service agencies, such as police and fire departments, and the reliability of their E911 services.
- How should the user interface into the converged network be considered?
- How much will the converged network cost, and which department will foot the bill?

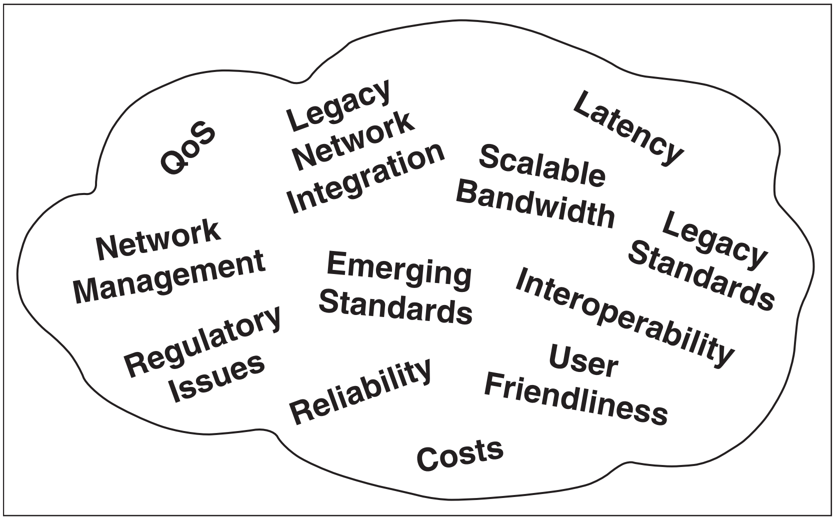


Figure 1-27: Challenges of the Converged Network

1.9 Looking Ahead

In this chapter, we have considered the principles of network convergence, the differences between voice and data networking infrastructures, the projections for growth in these respective areas, and the various players in this marketplace. Independent of these growth factors and which market research you choose to embrace, a general direction in today's networking environments points toward a convergence of the voice and data worlds—a trend that would be difficult to ignore.

The next and subsequent chapters drill down into these issues in greater detail, beginning in Chapter 2 with a discussion of the new applications that require a converged network infrastructure.

1.10 References

[1-1] The North American ISDN Users' Forum (NIUF) may be contacted at http://www.niuf.nist.gov.

[1-2] International Organization for Standardization. *Information Processing Systems—Open Systems Interconnection—Basic Reference Model*, ISO 7498-1984.

[1-3] Miller, Mark A. *Troubleshooting TCP/IP, 3rd Edition*. Foster City, CA: IDG Books Worldwide, Inc., 1999.

[1-4] Cavanagh, James P. "Defining the Three Markets of Telephony." Available from www.mbird.com/pdfs/wp_three_mkts.pdf, January 29, 2000.

[1-5] Fredj, Amy. "Interworking Between Next Generation IP and Legacy Networks." Available from www.commatch.com/data/websotefo;es/WP-Interworking Between NG IP and Legacy Networks.pdf, August 2000.

[1-6] Voice 2000: BCR's Guide to Emerging Voice Technologies. *Business Communications Review Supplement*, October 1998.

[1-7] Nicoll, Chris. "Multiservice Integration—Networks Converge on the Data Infrastructure." *Business Communications Review Supplement* (February 1999): 1–4.

[1-8] Miller, Mark A. "Voice, Video, and Fax over Internet Protocol: Possibilities for Converged Networks." *Decision Resources Spectrum Report*, April 15, 1999.

[1-9] Hu, Howard, and Houman Modarres. "IP Telephony—Paving the Way for Enhanced Services." Available from www.3com.com/other/pdfs/infra/corpinfo/en_us/50304601.pdf, Document number 503046-001, May 1999.

[1-10] Dickey, Clinton. "Voice and Data Convergence—Migration to an IP Telephony Network." Available from www.cirilium.com/gfx/news/ip_telephony_white_paper.doc, April 30, 2000.

[1-11] Covell, Andy. "Digital Convergence." *Network Computing* (December 11, 2000): 91–100.

[1-12] Richardson, Robert. "Unified Communications: Wireless, ASP, CPE Strategies." *Communications Convergence* (July 2001): 47–57.

[1-13] The number of Hosts and Servers on the Internet at any given time is calculated by Telcordia Technologies (formerly Bellcore) at http://www.netsizer.com/daily.html.

[1-14] The Yankee Group. "Global Network Strategies: The End User Speaks." 2001.

[1-15] Jackson, Edward R., and Andrew M. Schroepfer. *IP Telephony: Driving the Open Communications Revolution*. Piper Jaffray Equity Research Report, February 1999.

[1-16] Probe Research, Inc. *VoP Basic Services: Market Segmentation and Forecast*. CISS Bulletin, Volume 2, Number 4, 2001. For details, see www.proberesearch.com.

[1-17] Anderson, Rod. "Enhancing IP Network Performance: A Sensible Solution for Mission-Critical Applications." Nortel Networks Whitepaper WP503-2953EC-A, January 1998.

[1-18] Sage Research, Inc. *Primary Research Conducted in 2000*. For further information, see www.sageresearch.com.