Cross-Industry
Working Team

Building the Information Infrastructure: A Progress Report

Contents

July, 1999

Abstract

New communications transport and networking technologies are deployed in various combinations to build the communications networks for the national and global information infrastructure (NII/GII). Many of these technologies provide traditional communications services at lower cost, and enable new services that open new applications markets. This paper examines some of the new technologies and describes how they are being put to practical use. Examples of applications illustrate the essential differences among alternatives. The implications for existing services are described, and the impacts on progress to build the new information infrastructure are examined.

XIWT

The Cross-Industry Working Team (*XIWT*), is a membership organization of communications, computing, information and service providers working strategically together to advance the information infrastructure. *XIWT* publishes White Papers to create common understanding about technical, business and policy issues of relevance.

To obtain reports or for more information about *XIWT*, please contact:

Cross-Industry Working Team

Corporation for National Research Initiatives

1895 Preston White Drive, Suite 100

Reston, Virginia 20191-5434

Tel: 703-620-8990 Fax: 703-620-0913

Email: info-xiwt@cnri.reston.va.us

Web: http://www.xiwt.org

Charles N. Brownstein, Executive Director

Email: cbrownst@cnri.reston.va.us

1.0 Introduction

1.1 Background

XIWT's 1994 white paper, *An Architectural Framework for the National Information Infrastructure*, detailed the concept of "information infrastructure." It laid out an architectural framework for moving forward toward a new infrastructure capable of profoundly impacting the ways we work, learn, conduct commerce, do business and socialize. The paper set down the following guiding principles and specific goals for future infrastructure development. The new infrastructure should be widely available to all, operate in an open competitive economic environment, protect the rights of users and stakeholders, promote interoperability and open standards, and provide dependable high-quality services and an information marketplace.

A 1995 XIWT white paper, *Visions of the NII: Ten Scenarios*, described exemplary future uses of the infrastructure and the underlying technologies that must work together to make them possible. These creative speculations illustrated the possibilities and provided a basis for planning.

Since then, much has happened to bring the NII vision and ambition closer to reality, and to extend it globally. Public policies favoring less regulation of telecommunications provision and minimal regulation of information technology have become international goals intended to foster market competition, innovation, and economic growth. National governments began to deregulate telecommunications networks and to promote competition among industry segments. The World Trade Organization facilitated multilateral commitments among national governments to open their basic telecommunications services markets to competition. The Internet and IP-based technologies galvanized the marketplace. Many of the speculative technologies, services, and applications envisioned in XIWT's scenarios white paper are now available in the Internet marketplace.

The Internet has become the vision. It has transformed communications and business. In turn, it has been transformed from a research and education phenomenon into an expanding commercial public data network infrastructure, supporting ever-growing numbers of entrepreneurs, users and applications. It is *the* existence proof of great demand for an open, public data communications networking infrastructure along the NII/GII model. Its technical architecture, combined with a cooperative approach to development, operations, and governance, make the Internet today's premier incubator for innovations in computing and networking hardware, software, systems, services, and applications.

The growing use of the Internet has underscored a fundamental concern expressed in the 1994 XIWT white paper. The very flexibility and distributed operational characteristics of the Internet infrastructure that make it attractive also make it extraordinarily complex to understand and manage. This quality, in combination with rapid technology innovation and market growth, requires continual operational innovation and improvement to achieve the robust, reliable environment desired for economic and socially important applications.

In another white paper, *Class Profiles for the Current and Emerging NII*, XIWT identified a generic set of technology application profiles to associate information appliance and communications service functionality and performance levels with application requirements. The objective was to establish a framework for reaching a common understanding for application-oriented NII/GII requirements and to encourage industry members to set high functional and performance targets to satisfy user expectations. This reasoning is extended in this white paper.

Herein, we examine technologies that are presently being used to build communications networks for the new NII/GII. We examine the ways in which these technologies function, and the reasoning behind deployment decisions. There is a dual motivation for this analysis. The first is to assist non-engineers from information systems industries, public policy bodies, and regulatory agencies in understanding important new technologies that are influencing infrastructure directions and investment decisions. The second is to review progress toward achieving the closely coordinated bandwidth, response time, convenience, privacy, security, and interoperability characteristics desired for the new NII.

1.2 Analysis Approach and Organization

Information technology cannot be put to practical use in the abstract. It is brought to the market as hardware or software products, or network services. Today a diverse and growing array of hardware and software products is available for use in information networks. Making sense of all of the options can be a daunting task for network providers and application users alike. This paper introduces a “what and where” model to facilitate the examination and analysis of how multiple networking products are combined to support valuable end-to-end applications. This model, described in detail in section **2.0 Framework Model**, associates the *functions* required to make a networked application work with specific *technologies* in the deployed distributed networking environment.

The technology topics discussed in the remainder of the paper were selected to address the status of NII/GII development, trends, implications, and open issues. These are as follows:

3.0 Access Technologies — network value increases as more users have access to network services; cost-effective access at higher bit rates is a major technological challenge.

4.0 Interworking — support for disparate networks, systems, services, and applications is essential to achieving a transparent “out of sight, out of mind” infrastructure.

5.0 Support for Legacy Systems on the New NII — there are few economically viable opportunities for “green field” starts for broadly deployable networks.

6.0 Voice over Packet-Switched Networks — integrated network services can realize the full potential of networked applications.

7.0 Optical Transport Technology — affordable bandwidth is a key to continued NII innovation and market growth.

8.0 Middleware — a critical concept for achieving open systems and interoperability is common software for multiple applications.

The guiding principles and goals outlined in *An Architectural Framework for the National Information Infrastructure* provide the basis for judging progress toward the new NII/GII in **9.0 Status and Outlook**.

2.0 Framework Model

This paper introduces a "what and where" matrix modeling method to depict how computing and networking functions are distributed across interconnected network components to implement useful services and applications. The reference model matrix, described below, is used throughout the paper to explain the interoperation of a mix of technologies including protocols, physical equipment, software, and network service offerings, in specific application examples.

2.1 Reference Model Matrix Overview

The reference model matrix combines the ISO standard for open system interconnection (OSI), defining a framework for implementing communications protocols in seven layers, with a second axis that depicts the geographic distribution of network functions in realizable technology components. Technology components are devices bought and deployed by individual end-users, organizations, and communications network operating companies (network carriers).

Figure 2-1 is a simplified representation of the physical layer of an end-to-end information network. It is used here to depict the seven geographic network segments used in the reference model matrix, consisting of: information appliances; the major autonomous networking domains (user/enterprise premises networks, access networks, and backbone networks); and the network hubs where switching and routing systems are used to support inter-network domain connections. These segments are defined as follows:

Information appliances — PCs, workstations, and other devices (e.g., telephone instruments, fax machines) that are connected to the network.

Premises networks — customer-owned premises-based communications networks, e.g., local area networks (LANs) and premises wiring.

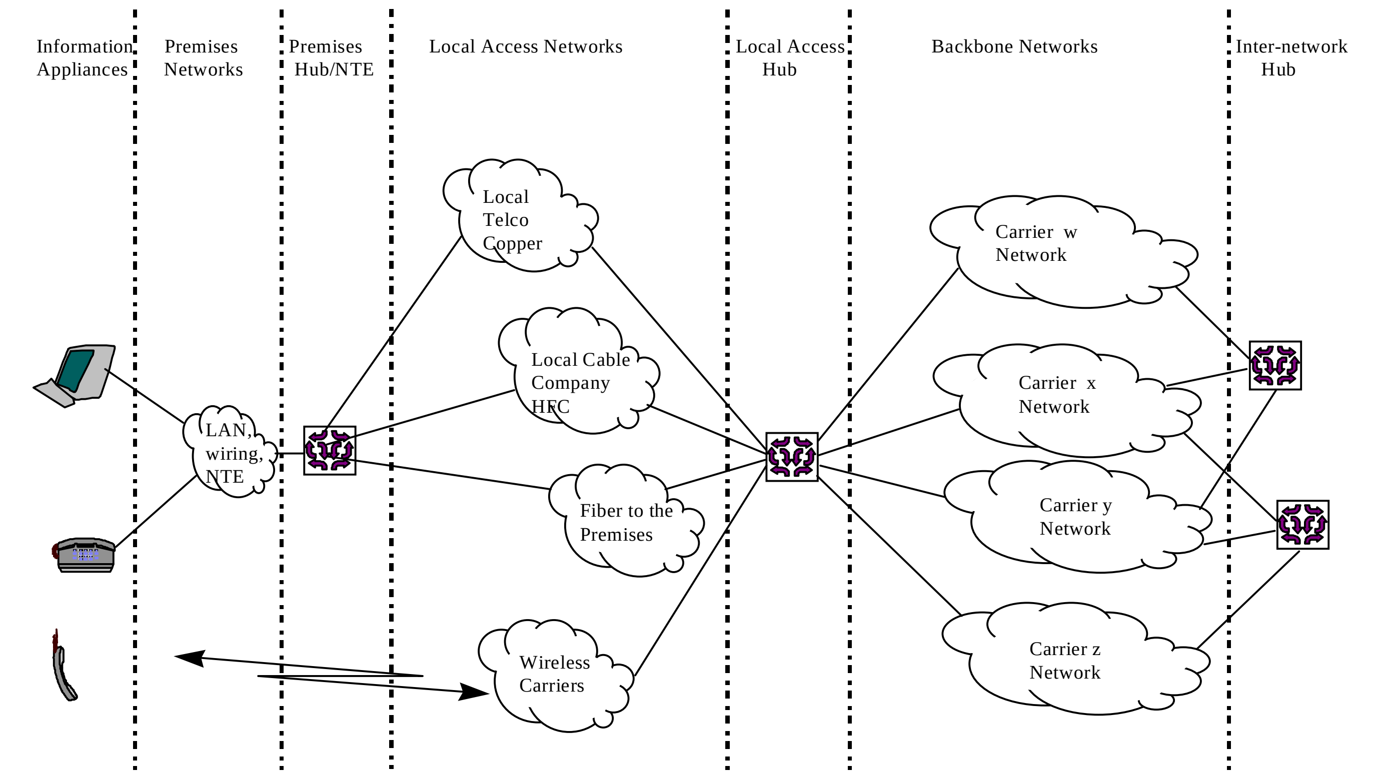

Figure 2-1: Geographically Distributed Network Segments

Premises hubs/network terminating equipment (NTE) — equipment that provides premises networks with access to network interconnection functions such as signal conversion, traffic aggregation/grooming, routing, and switching. Examples of such equipment include access multiplexers, edge switches, routers, PBX, modems, and customer service units (CSU).

Local access networks — the network facilities and services provided by local access carriers including local telephone companies, cable companies, wireless carriers, and alternative access carriers.

Local access hubs — switching and routing centers located at the edge of the access networks where customer information is switched and routed to regional/national/international backbone networks to reach destinations defined by customer-initiated signaling. These may provide inter-service gateway functions.

Backbone networks — the network facilities and services (transport and switching) provided by regional, national, and international carriers.

Inter-network hub — the switching and routing centers for interconnecting backbone network facilities and services. These may provide inter-service gateway functions.

Note that the figure does not show the numerous computers providing higher level functions within the network segments; the location and operation of these are depicted in the full reference model matrix (figure 2-2). This matrix is populated with technology components that can be bought and deployed in the form of hardware, software, and services.

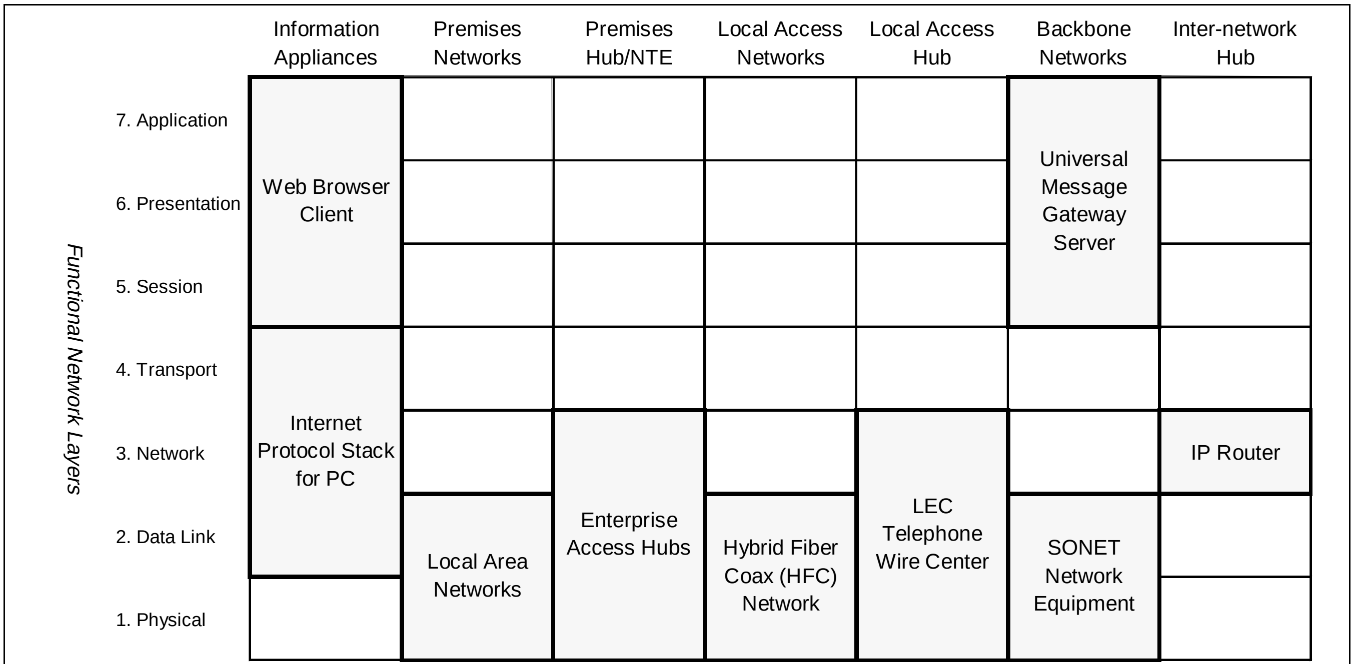

Figure 2-2: Reference Model Matrix

Functional Network Layers

	Protocols	
7. Application	E-Mail, Web Browser, FTP, Telnet	Internet Protocol Suite (IP) enabled applications utilizing the functions and services of the underlying networking layers.
6. Presentation	HTML, GIF, MPEG, JPEG, WAV	Numerous standardized formats for data coding and data interchange independent of the application and specific vendor. In addition there are coding formats associated with specific applications environments .
5. Session	<Null>	There is no functional equivalent to session in the Internet Protocols.
4. Transport	TCP, HTTP, UDP, ICMP	Transmission Control Protocol (TCP) is the Internet layer 4 transport standard, it provides a reliable, sequenced, flow-oriented end-to-end data stream; Hypertext Transport Protocol (http) is a transfer protocol used by the World hypermedia system to retrieve distributed objects. HTTP uses TCP as a transport layer. User Datagram Protocol extension to IP that provides multiplexing of IP packets between pairs of hosts. Internet Control Message Protocol error reporting and control mechanism. Both UDP and ICMP
3. Network	IP, routing protocols	Internet Protocol (IP) routes and delivers information between pairs of endpoints and through the intermediate network nodes. IP provides a service which is connectionless, best effort and unreliable. IP packets may be lost, duplicated, delayed and delivered out-of-order.
2. Data Link	PPP, 802.n	The IP protocol runs over a data link in an underlying subnetwork. PPP is the most common protocol used by users with modem access to an ISP, while 802.n protocols are common in the enterprise LAN environment. ATM, Frame Relay, X.25 and SMDS are used in local and backbone carrier networks
1. Physical	V.x, 802.n, DSn, SONET	Various physical networks and subnetworks are used to provide the digital bitways V.x refers to modem standards and 802.n to LAN standards. DSn and SONET are standards for local and backbone carrier networks.

Figure 2-3: Example of a Technology Baseline Chart: Internet Technology

2.2 Using the Reference Model Concept

The reference model matrix is used to show how technology is packaged and deployed in technology components, and how these components are interconnected to support useful services and applications.

The first step in using the model is to establish accurate functional baselines for the technologies of interest. This is done by mapping the protocols and standards associated with each technology to the seven OSI model functional layers (i.e., physical, data link, network, transport, session, presentation, and application). Although most of the functionality defined in the OSI model is present in all communications systems, few are implemented precisely according to the OSI model. It is not a problem if the functional coverage of the protocols and standards that define a specific technology are not perfectly aligned with the OSI layer boundaries. Some layers may be empty, or two or three layers may be implemented in one. The Internet technology baseline chart, mapping Internet protocols and standards to the OSI layers, is shown in figure 2-3.

The next step is to describe how each technology is put to work in the form of practical technology components that can be bought and deployed. The functional coverage of each component is mapped to the OSI model functional layers as shown in figure 2-4.

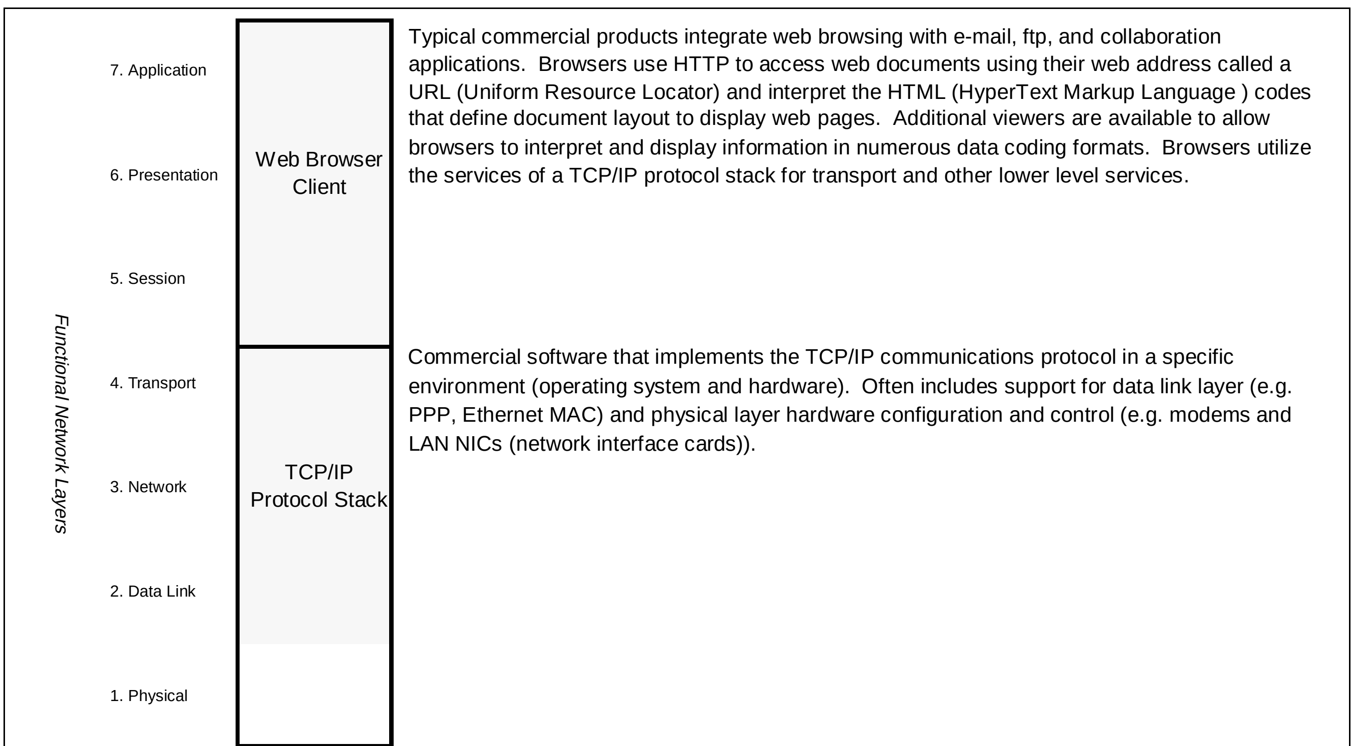

Figure 2-4: Example of Component Functional Coverage

The reference model matrix is used to show how these components are deployed in practical networks. Depending on its functional coverage, a given component may be deployed in one or more of the network segments as illustrated in figure 2-5.

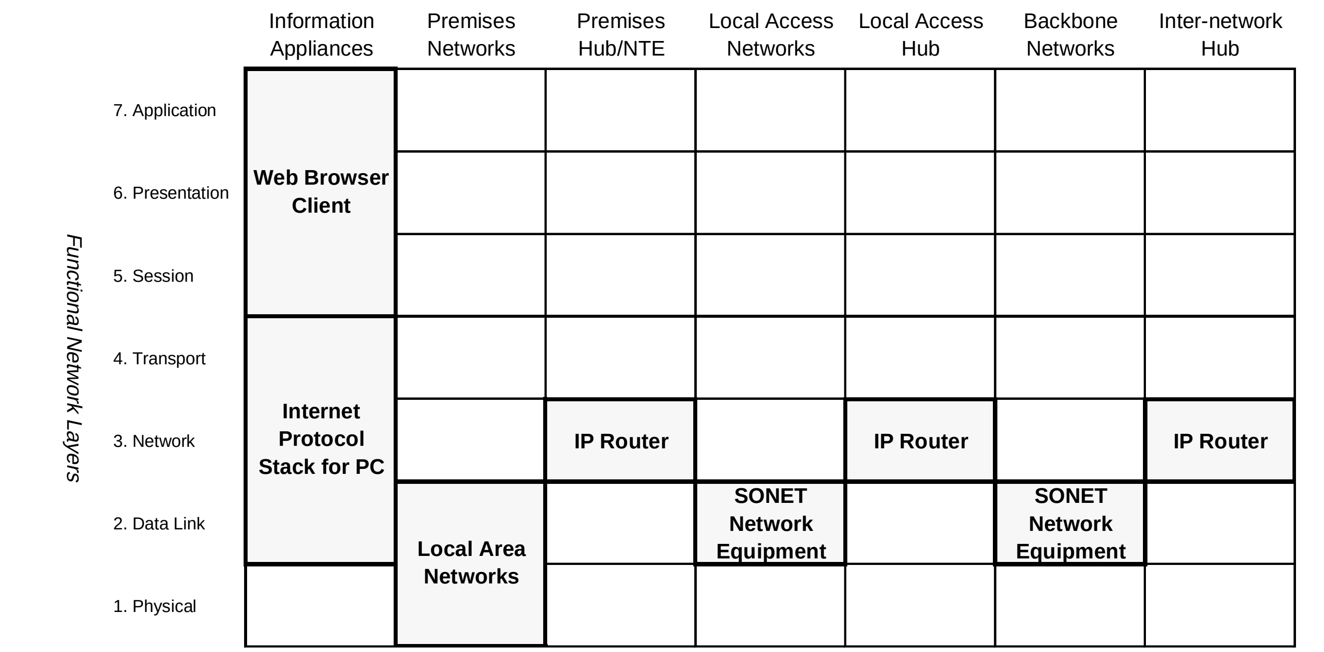

Figure 2-5: Mapping Technology Components

In the final step, an end-to-end version of the reference model matrix (for the case of dialup access to a corporate network) provides an example of how these components are deployed to support specific networking service and application examples (figure 2-6). More often than not, several technologies are involved. The reference model matrix depiction shows how practical networks are interconnected, and identifies the interworking requirements that must be met.

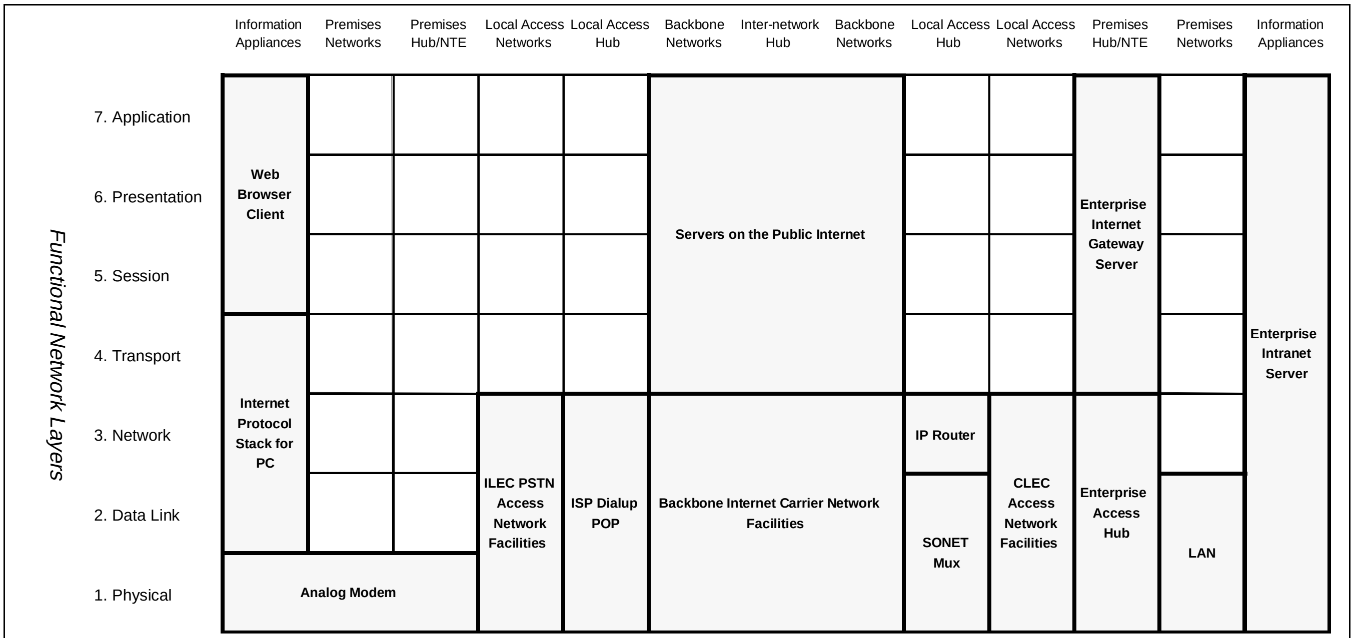

Figure 2-6: Dialup Access to Corporate Network Example

3.0 Access Technologies

The value of a communications service increases with the number of locations it serves and the number of individuals who use it. **Backbone networks** are deployed to interconnect geographically dispersed locations. **Local access networks** are deployed to connect individual users to backbone network services. This section examines the present state of and future trends for local access networks.

Developing a high-performance integrated information infrastructure for the information age — the new NII/GII — is a national goal of the United States. A major objective of the Telecommunications Act of 1996 is to promote a competitive environment in which old and new communications providers build a network of interconnected networks that supports new interactive multimedia services on a widespread basis. Figure 3-1 is a conceptual model of this new networked multimedia environment. It supports networked applications with a heterogeneous infrastructure of multimode information appliances and interconnected communications networks.

3.1 Embedded Networks Are Purpose Built

Historically, communications networks were built to provide a specific service. Little attention was paid at the time of their construction to issues of integrating future services over a common network infrastructure. Public telephone networks, including wireless ones, were designed and deployed to handle voice traffic. Similarly, cable television networks were optimized for one-way video broadcasting, and the Internet was initially developed and optimized for best-effort, store-and-forward, packet data transport. This "service-specific optimization" is most prevalent in local access networks, where deployment and operating costs are directly associated with individual customers.

Nevertheless, tremendous customer interest has lead to widespread use of the public telephone network to gain access to the public Internet. Analog modem technology, which makes data appear like voice signals to connected equipment, has made the telephone system today's most ubiquitous data access network, at speeds up to 56 Kbps. Figure 2-6 illustrates how the public switched telephone network (PSTN) is used to provide access to the Internet. In this example, remote users gain access to corporate applications and information through the public Internet. The corporate network connects to the Internet using access services provided by a competitive local exchange carrier (CLEC) specializing in high-speed data access services. The remote user, equipped with an Internet-enabled PC and a dial-up modem, makes a local telephone call through the local telephone company's network (incumbent local exchange company or ILEC) to connect to the Internet service provider's (ISP's) local point of presence (POP). The POP contains a shared modem pool, routers, and communications equipment to allow customer traffic to be consolidated and routed through the backbone Internet carrier networks and the CLEC facilities to the corporate network. In the example, a corporate "firewall" provides security and user authentication functions to protect the corporate network from unauthorized access from the public Internet.

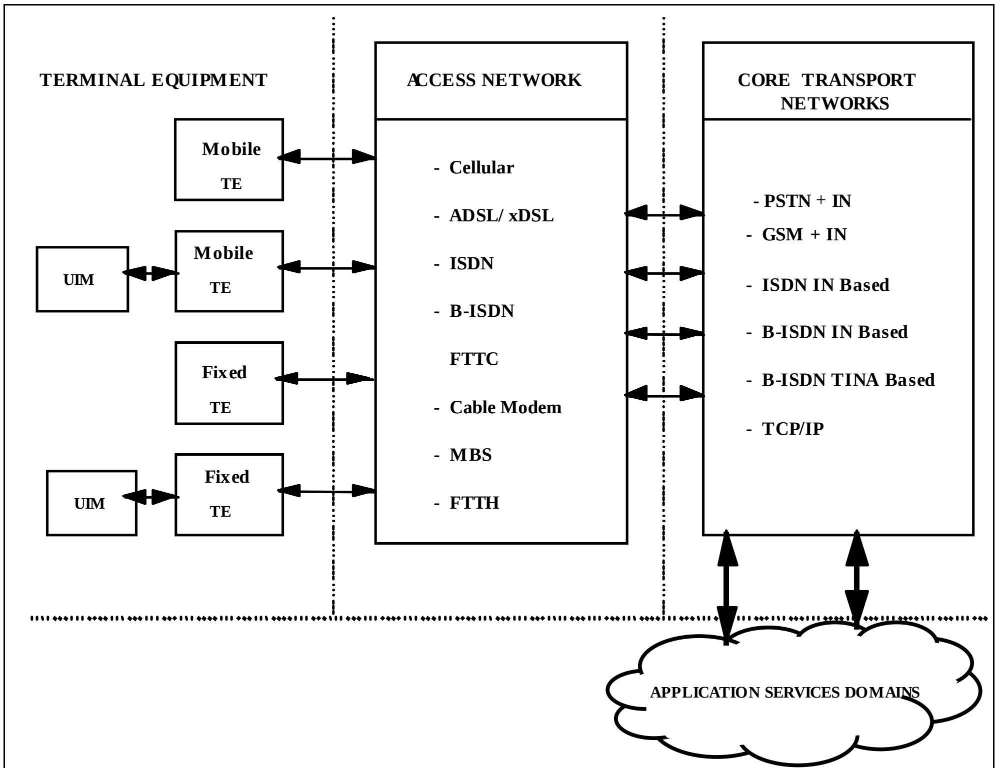

Figure 3-1 Multimedia Communications Technology Environment

3.2 New Market Forces at Work

The growing utility and popularity of electronic commerce and of interactive multimedia Internet applications have created a requirement to integrate voice, video, and data. With experience, users are developing sophisticated demands regarding the performance levels network services must provide to satisfy their bandwidth-hungry applications. These applications severely tax modem access technology as it has evolved through the PSTN. The result is frustratingly long latencies — the notorious “world wide wait” for the large amounts of information to return from requests made by users of the World Wide Web, or impaired sound quality for Internet telephony or audio conferencing.

Communications service providers recognize that their future success depends on how well they respond to rapidly changing service demands. They understand that their service-specific networks can barely support integrated interactive services today, let alone respond to projected levels of demand. Providers have thus begun to make revolutionary changes in service models and enormous investments in new network facilities to remain viable in the new competitive environment. Perhaps the most crucial issue for providers is how to approach modernization of their embedded access networks to support new services and applications economically.

3.3 Provider Modernization Strategies

High-speed connectivity to homes, small businesses, and remote offices creates many new opportunities for service providers. These include high-speed Internet connections that are always connected (always-on), on-demand and interactive multimedia entertainment, and peer-to-peer video collaboration and multi-player game applications. Connections at rates of one Mbps or higher are required to enjoy the full capabilities and benefits of such multimedia applications. However, it is not yet economically feasible for the incumbent carriers or new competitors to fully replace access networks. Instead, these local telephone and cable companies have developed various improvement strategies to modernize existing access networks.

Many local telephone companies are beginning to deploy one or more versions of digital subscriber line technology (xDSL) in their access networks. The most important feature of DSL is that it can overlay high-speed digital services on the existing twisted pair copper network, without interfering with the traditional analog plain old telephone service (POTS). DSL allows subscribers to retain the services to which they have already subscribed and to add new services such as high-speed Internet access and, with some versions of DSL, digital video on demand. Subscribers install a compatible DSL modem to send and receive the data that comprise these services. DSL works by using sophisticated digital signal processing to gain access to the frequency spectrum above that which is reserved for voice. Most DSL implementations will be asymmetric (ADSL), with the higher data rate downstream from network to subscriber.

Cable operators are upgrading their networks with fiber optics in the feeder plant to support two-way traffic to neighborhood distribution hubs. The final connections to the subscriber are through shortened coaxial cable distribution, with cables passing between 100 and 500 homes. This configuration is known as hybrid fiber coax (HFC). One or more TV channel slots on the cable are dedicated to high-speed data access services.

Subscribers purchase or rent a compatible cable modem to send and receive data over these channels.

These two emerging access technologies promise to provide individual users with high-speed data network access at a low cost. Figure 3-2 illustrates how these technologies are deployed to support high-speed access. Cable modems are capable of higher data rates than DSL, but this advantage may be lost under actual usage conditions. Cable modems share the access medium in such a way that connected devices share and thus may compete with each other for the available network bandwidth. The actual achieved bandwidth at any given time varies with the number of simultaneous users. DSL access lines are dedicated to an individual subscriber. The customer's DSL modem is linked via the PSTN to a dedicated line card at the providing telephone company's central office facility. The various types of DSL provide a wide range of data rates. Additional variability occurs due the length of the copper access lines (distance from modem to line card).

Modernization plans for wireless services include introducing higher data rates in mobile cellular access networks and deployment of new satellite based systems to offer worldwide mobile access to broadband services. These satellite systems use a constellation of low-orbit satellites to provide worldwide access at data rates from 64 Kbps to 60 Mbps. Services offered by satellite access include mobile voice, Internet high-speed access, messaging, paging services and broadband services.

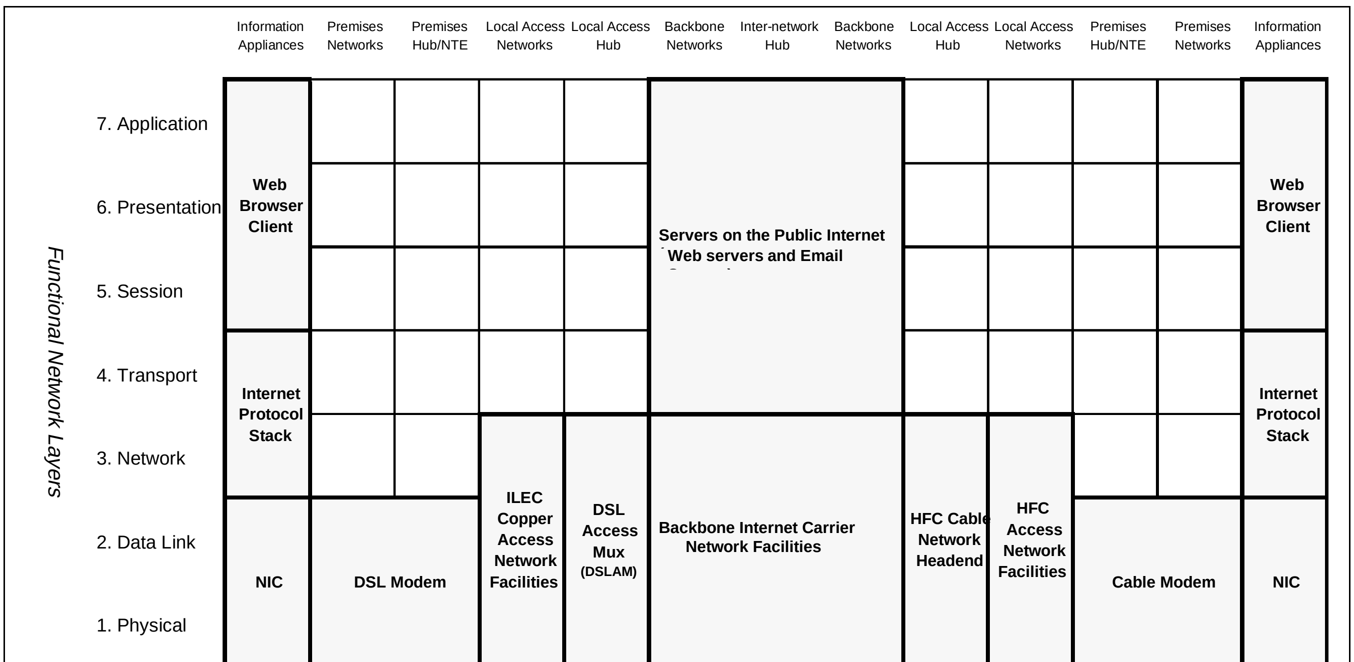

Figure 3-2: DSL and Cable Modem Access Network Configurations

Wireless technology (MMDS, LMDS) is also being considered as an alternative to wired networks to provide broadband access to users at fixed locations. Wireless data access networks employ radio transmission technology in different portions of the radio spectrum, but they all have a similar structure. Figure 3-3 depicts a mobile user accessing a corporate network over a wireless data connection. User terminals are equipped with a radio transceiver/data modem and an IP stack for mobile operation. The remote terminal communicates with a local access base station over a radio link. A mobile access POP supports interconnection with backbone network services.

These new access technologies by themselves cannot guarantee high performance. Application functionality and performance are determined by how well the information appliances, servers, and communications services are matched on an end-to-end basis. For additional information on this topic, see the XIWT white papers, *Class Profiles for the Current and Emerging NII* (1997) and *Customer View of Internet Service Performance* (1998).

3.4 Future Directions: Toward Integrated Services

Many core network carriers are adopting asynchronous transfer mode (ATM) technology for their backbone network upgrades because ATM promises true multiservice capability. ATM makes effective use of network resources by providing bandwidth on demand and sharing bandwidth among parallel applications. ATM supports multiple quality of service (QoS) classes including the most demanding multimedia applications. Its proponents characterize ATM as an “integrator technology” which can consolidate bursty data traffic with constant rate analog real-time voice and video over a single manageable infrastructure.

ATM is a cell-switched, connection-oriented technology that uses switches to establish virtual connections from an input port of an appliance to an output port of another appliance. It uses 53-byte, fixed-length cells that can be switched at very high speed. ATM supports both switched virtual connections, for service similar to the voice telephone network, and permanent virtual connections to support private line services.

ATM is also offered in access networks to provide end-to-end service interoperability with core network services. It is being used with xDSL to allow simultaneous connectivity to multiple services, multi-protocol support, and multiple services classes with QoS support. ATM also provides a migration path for evolving access technologies. Providers of wireless broadband networks plan to use ATM to facilitate the use of high network bandwidth.

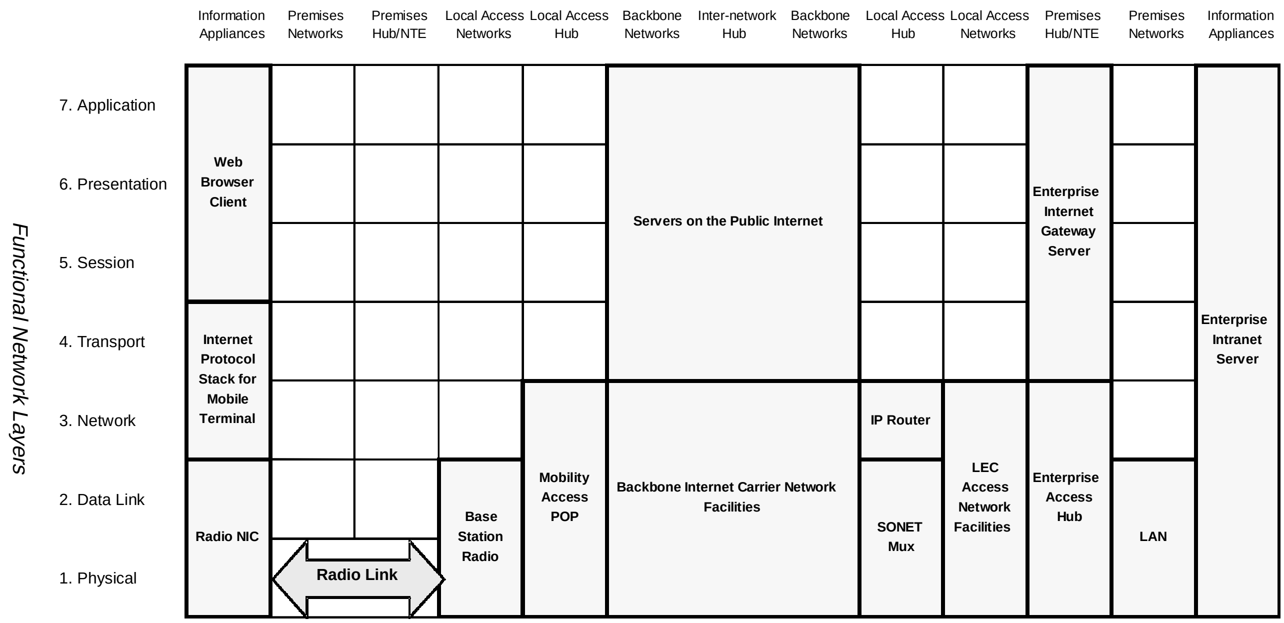

Figure 3-3: Wireless Access to Corporate Data Network

Over time, ever-increasing bandwidth requirements will extend fiber optics technologies from the backbone level, where they are presently employed, into access networks closer to the subscriber premises. Integrated access architectures are being proposed that will provide a full range of “future-safe” ultra-high-capacity services (e.g., with excess capacity over foreseeable needs) while maintaining compatibility with present network facilities. These new systems will allow carriers to make capital investments in access networks as new markets develop. The multiservice access network architecture depicted in figure 3-4 is analogous to both the digital loop carrier configurations used by telephone companies and the HFC configurations used by the cable industry. It supports a wide range of traditional narrowband interfaces and services for residential and small business subscribers. However, its real strength is an ability to implement emerging broadband services including switched broadband and digital video on demand.

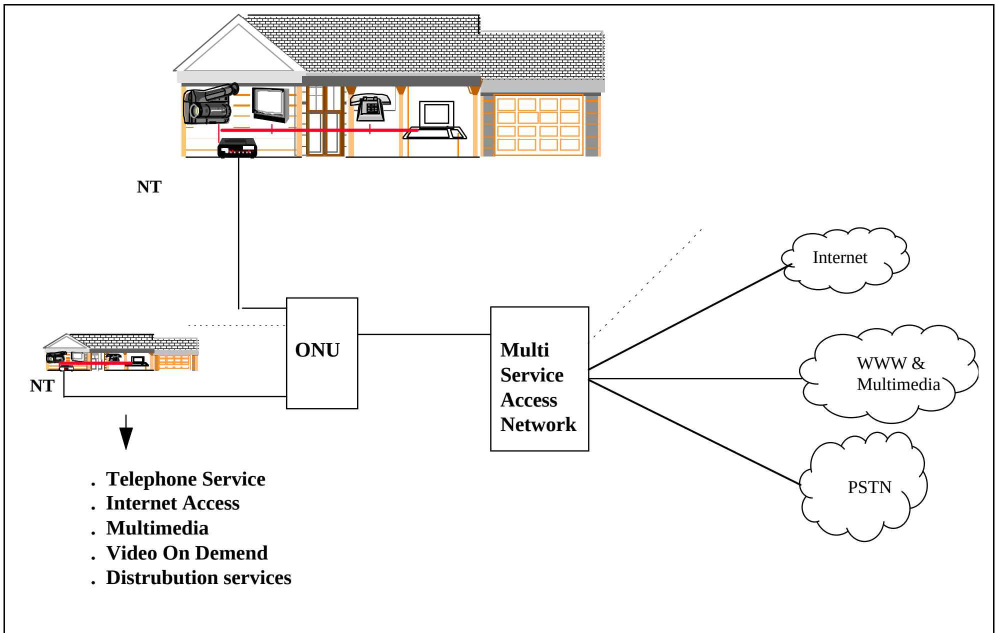

Figure 3-4 Multiservice Access Network Architecture

The multiservice access node located at the telephone wire center or cable head-end provides access to core network services. It uses ATM to multiplex multiservice subscriber traffic for transport over a fiber optic link to the remote optical network unit (ONU). The ONU terminates the fiber and separates the signals for distribution to a cluster of 10 to 200 nearby subscribers. The ONU is equipped with a mixture of interfaces supporting subscriber access over copper pairs and coax to maintain full compatibility with analog telephones, DSL modems, televisions, and cable modems. It also has sufficient flexibility to support subscriber access over fiber optics when demand develops.

4.0 Interworking

One common thread theme in virtually all visions of the NII/GII is that end-users must have access to services in a simple and efficient manner. Moreover, the service should provide the customer with the performance agreed to at the time of purchase. This implies a high, yet unattained, level of cooperation and integration among the various networks and/or service suppliers. Achieving this seemingly simple goal requires some extraordinarily complex technical and business cooperation, or interworking.

Today's information infrastructure is a patchwork of heterogeneous networks, including public switched telephone networks, private custom data networks, the public Internet and specialized subnets, or intranets. Beyond announcements of new "whiz-bang" technologies, products, and services, lies the vast challenge of interworking*: How do we make it all work together?* Interworking is a complex problem, but is critical to the success of the NII/GII.

For the sake of simplicity, we focus on three aspects of interworking:

- network-to-network,
- service and provisioning
- Quality of Service and operational interworking.

While it may be easier to speak of these as if they are stand-alone items, the fact is that effective interworking assumes a complex interdependency among these capabilities (and much more).

4.1 Network-to-Network Interworking

Perhaps the most obvious form of interworking is that which is required between networks to complete a connection for an end-user. One of the more obvious examples of this kind of interworking takes place when a voice telephone call is handed off from a local exchange carrier (LEC), to an inter-exchange carrier (IXC), and ultimately back to another LEC. For this relatively “simple” example to work, the three networks involved must be able to communicate over a signaling network. The equipment that each network deploys must be capable of working with the equipment deployed by the next carrier in the chain. Standard protocols have been developed to allow this type of interconnection. This very basic type of voice connection continues to represent a large portion of connections placed today. But this simplistic call model is being replaced by more complex uses of the NII.

For example, consider the following scenario in which only today’s technologies are used. Let us assume that one is on a business trip. After the day’s work, they want to call into their home office to download a file from their desktop PC. When they get to their hotel room in the evening, they plug a laptop into the telephone jack in ther room and dial up the local POP to connect to their ISP. Through the ISP’s network, they will eventually link to their corporate LAN.

This scenario also involves quite a few networks. Figure 4-1 provides a quick overview of the relationships that might exist between networks at some point during the connection.

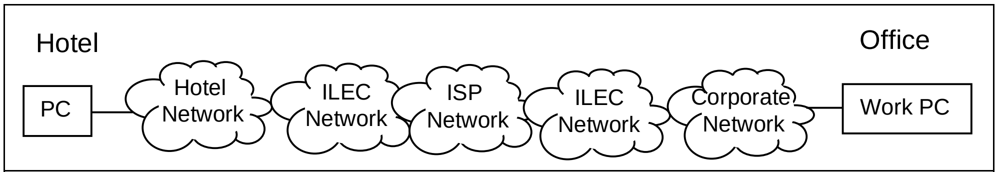

Figure 4-1 Interworking through Multiple Networks

Most of those networks will be transparent. When one plugs their laptop into the hotel’s telephone jack, they know its modem must be configured to dial a “9” to obtain an outside line. Beyond that, however, they need know little else (if anything) about the hotel’s network. The ILEC network is apparent only in that they have to know the local telephone number of their ISP. The ISP network is apparent only to the degree that some type of “login” process is required. Once connected to the ISP network, they log onto their corporate network. At this point, they are unaware of any of the details about the connections between her corporate network and the ISP network in her hometown. Clearly, each of these networks must work effectively and efficiently to complete the connection to the corporate computer.

The networks shown as clouds in figure 4-1 could easily be expanded to show an incredibly complex range of detail at the physical level. The ISP "network" may, in fact, be a "virtual" construct using elements from many networks. For example, the ISP's network may utilize high-speed circuits leased from one IXC between the travel city and the ISP home location. There, technologies may be employed that could use a different carrier's backbone network to connect to the ISP's site in the hometown. Conceivably, portions of the data communicated (segmented into ordered data packets) could traverse different paths between the hotel and her hometown. A call may begin with the packets routed through the travel city, but at various times during its life, packets might be routed through other cities. These packets would be properly reassembled in the computer at the destination for proper use by the application software.

The interworking of physical devices requires another component, protocols. Virtually all communications in the future NII will be based on protocols and interworking among these protocols. For simplicity, we use the seven-layer OSI model in this discussion. Figure 4-2 shows the same high-level network sketch depicted in figure 4-1, but with the addition of a "protocol stack" above each network. The intent here is to indicate that each network may operate within the OSI stack, but it is *not* essential (and may not even be possible) for each network to operate at *all* layers of the OSI stack. There are several principles at work here.

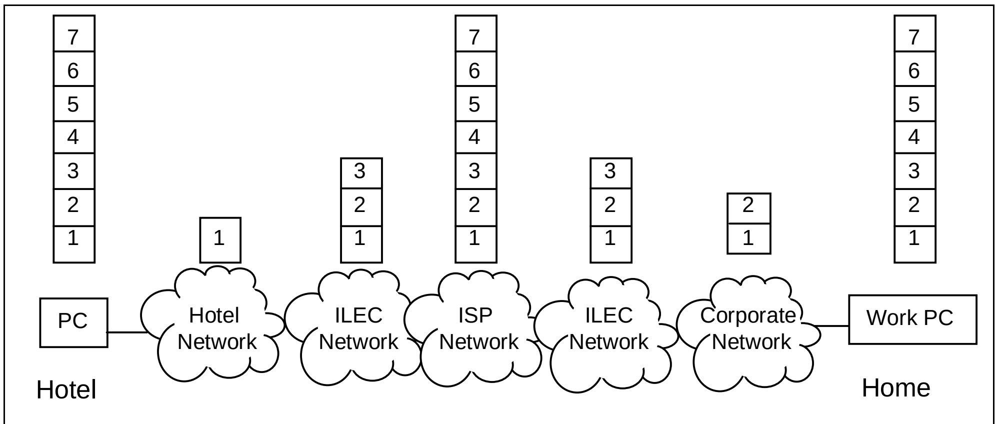

Figure 4-2 Functions Will Vary across an End-to-End Network

One very important concept is that no network, or network component, should operate at a higher level in the protocol stack than is necessary. Another concept is that virtually every network will utilize a variety of protocols. For example, figure 4-3 shows that a variety of protocols will be used in virtually *every* network along the path of Beth's call. Thus, the protocols operating at layer 2 of any of the network components shown below may *not* necessarily be operating in other devices at their respective layer 2 elements.

Clearly, for the connection to work per Beth's expectations, these networks must be able to communicate with each other, matching speeds, equipment signal formats, and a variety of other requirements to complete the call. The networks must *interwork*.

In terms of physical interworking then, it should be clear that today's networks already pose a very complex problem for efficient networking. Figure 4-3 hints at the level of complexity involved. Some of these protocols originated from within the public network domain, while others originated from within the computing industry. For these networks to interoperate effectively, a varied and complex set of protocols has been developed

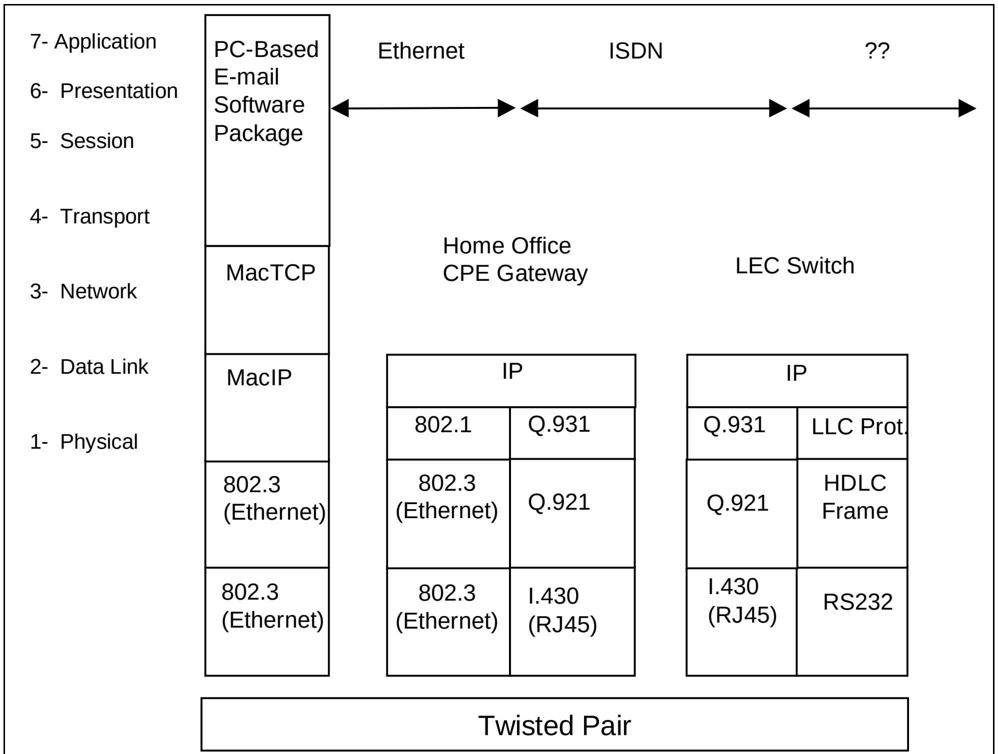

Figure 4-3 Example Protocols

4.2 Service and Provisioning Interworking

The complexity of the physical network outlined above suggests another interworking concern, service and provisioning transparency. When Beth initiated her session, she was not aware of how many network segments (or networks) would be involved, or what intermediate bit rates or protocols might be employed by the various involved carriers. This *transparency* of the networks to the end-user is absolutely critical to the NII. Knowledge of how connections are provisioned will normally not be important or interesting to end-users. Service providers will create "end-to-end" solutions for customers. For maximum efficiency, the marketplace should determine the relationships that carriers can form with each other to provide such solutions.

If providing the basic connectivity described above seems complex, the provisioning of future value-added services will be even more so. Hints of this emerging complexity already exist as new services emerge. One is Internet call waiting (ICW). In its simplest form, the service would notify an Internet user, via a message to the user's computer screen, of an incoming call on the line being used for the Internet connection. The user would be given several options for how to handle the call. That is the call could be rejected (with the caller receiving a busy signal); forwarded to a voice mail system; or accepted, in which case the Internet session would be gracefully terminated to make way for the incoming call. Although, to the user, this service has characteristics similar to those of call waiting on the voice network, it is technologically quite a bit different and will pose a number of interworking problems. Obviously, the PSTN switch will need to interwork with the ISP network router. This suggests a need for innovative and non-standard signaling and call control mechanisms.

Moreover, just as the signaling and call control issues may need to be worked out for physical interconnection, another set of issues needs to be addressed in terms of service provision. For example, to provide ICW service to the end-user, elements of both the PSTN and ISP networks must work together. In this case, from the end-user's perspective, who will be the service provider? How will billing work? Whom does the customer call when there is a problem with service?

ICW is a relatively simple service. Interworking issues in terms of service provision and support can and will become overwhelming in the future, particularly for more complex future services. ICW could (and likely will) be offered soon by carriers who can provide both PSTN and ISP functionality. When we picture a world in which such services may be subscribed on a call-by-call basis (e.g., a customer subscribes on a "one time" basis because they are expecting an important call), the complexity increases dramatically.

Similar complexity can be envisioned for the various electronic commerce services, information services, and other capabilities envisioned over the long term for the NII. All of these must be offered in a manner that is efficient and easy for the end-user. It is thus clear that the future NII will require intense interworking not only at the physical networking level, but also in terms of services and service provisioning.

4.3 QoS and Operational Interworking

As networking becomes more complex and dynamic, the challenges of providing a given quality of service also increase. For example, suppose that Beth's application required that the connection from hotel to home office be at a particular speed (perhaps 3.2 Mbps), and that the bits were to be delivered in sequence, with some minimum latency (delay) in their transit. For the application to perform as expected, each intermediate networking provider involved in creating the circuit must now "negotiate" with the one before and the one after it in the chain, and agree to accept the customer's conditions. This interworking is taking place at the level of quality of service.

Clearly, some kind of signaling or equivalent will be required between these networks, both at setup and whenever a network wants to make a change during the life of the session. This kind of signaling has been defined for ATM networks, and is well under way for IP networks. However, developing effective interfaces to ensure QoS across any and all networks is quite another story. A great deal of research and development is needed to provide the kind of QoS customers will expect in the future NII.

The issues are not solely technical. Suppose Beth requests a certain set of QoS settings for her call. In today's environment, the LEC would normally link Beth to her predetermined IXC for transport across Local Access Transport Area (LATA) boundaries. What happens, however, when the customer's IXC cannot provide the requisite QoS settings at that point in time? Normally, the call would be rejected by the network, and the customer's CPE would make another attempt at a less rigorous QoS. This process continues until a group of settings can be met by all networks in the chain. What if the first network in the chain were not capable of negotiating with a variety of other networks? While the customer's pre-selected IXC might not be able to handle the call at a given setting, perhaps another network might exist that *could* supply the specified connection setting. Could the first network in the chain offer the customer the choice of either relaxing the call's constraints or using a different carrier? This type of customer interface would certainly seem possible in the future. However, implementing it would involve public policy as well as technology.

5.0 Support for Legacy Systems on the New NII

IP and ATM are emerging as the core networking technologies for the new infrastructure. Taken together, these technologies exhibit many of the characteristics needed to support demanding interactive networked multimedia applications. However, good support for emerging applications is not enough. The pace and scope of deployment of new networking infrastructure also depends on how well the new structures and technologies accommodate the transition from embedded legacy systems. It is not reasonable to expect businesses simply to discard such systems. The cost would be too high, and replacements cannot be implemented overnight. Many legacy systems efficiently support the mission-critical processes for which they were developed. A very real issue facing industry today is how best to consolidate and integrate valuable legacy systems with the new information infrastructure environment of global interactive networking. Three primary integration/consolidation needs must be addressed to deal with this dilemma successfully.

- Consolidation of network traffic,
- Connection of legacy networks, and
- Access to legacy applications and data.

5.1 Cost-Effective Network Traffic Consolidation

Businesses depend on their internal mission-critical legacy information applications, but also understand the need to develop new applications for the more open-networked business model of the future. At the same time, costs must be controlled if businesses are to remain competitive. Using a single networking approach to support both legacy and new applications is a strategy both to manage recurring communications costs and to build an infrastructure for the future.

Today's legacy applications often run over separate private networks that use proprietary protocols, e.g., SNA, IPX, DECnet, or AppleTalk. Legacy networks are designed to achieve high levels of stability and robustness, and consolidated network operation must maintain these legacy service levels.

ATM is specifically designed to support a mixture of traffic types in a connection-oriented network environment. ATM not only can provide the increased bandwidth needed in the new applications environment, but also is designed to support multiservice traffic so that voice, video, and data can be consolidated (figure 5-1).

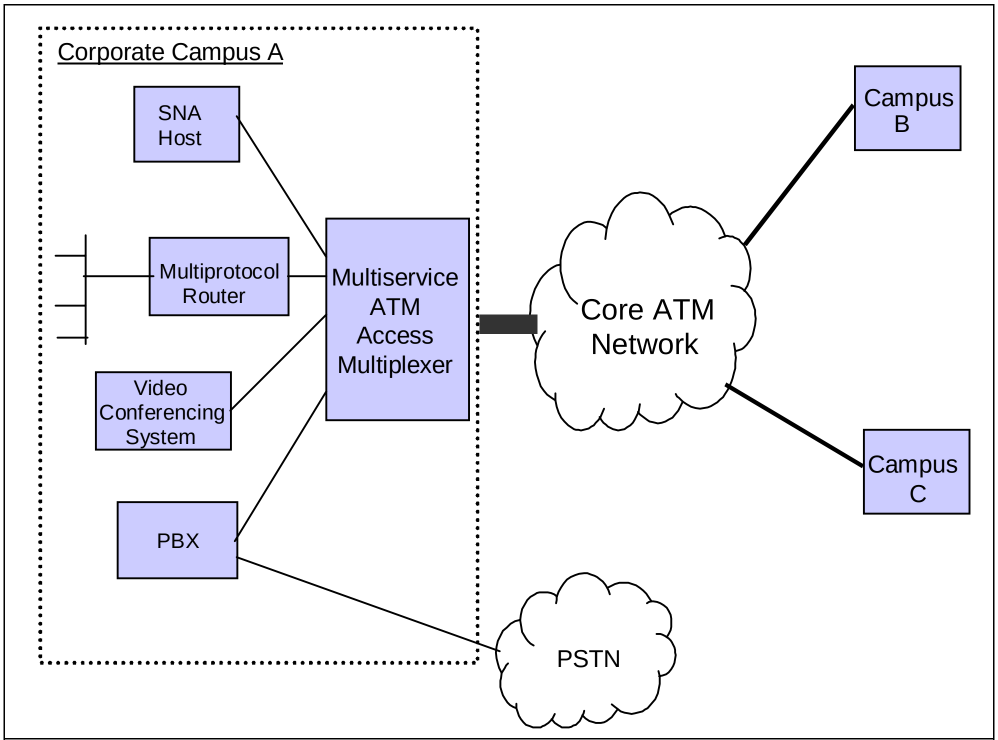

Figure 5-1: Corporate Wide Area Network Consolidation

The QoS provided by ATM supports applications with adequate and consistent delivery of information. In addition to being a core technology in backbone networks, ATM is used in campus networks to support high-speed requirements. LAN emulation and multiprotocol over ATM (MPOA) are specifications set forth by the ATM Forum standards body that allow non-ATM packet-switched protocols (IP, IPX, etc.) to be routed over ATM networks. MPOA uses the concept of a virtual router — an MPOA route server — that does routing calculations for the target protocols and uses the results to establish virtual connections through the ATM network. The next hop routing protocol provides additional capability for direct connections over ATM.

Multiprotocol routers are the primary internetworking tool in connectionless packet networks. They support two or more communications protocols, such as IPX, SNA, TCP/IP, and DECnet. They are used to route packet traffic between different LANs located throughout an enterprise. They can also interconnect LAN traffic to wide area network (WAN) facilities such as private lines and other switched network services, including ATM. Routers can also manage link bandwidth. For example, when SNA is mixed with other protocols, the traffic can be managed to guarantee SNA response time. Routers can also minimize broadcast traffic, especially over the WAN where recurring costs are a concern. A router can secure a network by filtering traffic from unauthorized users based upon OSI layer 3, 4, or higher information.

Multiprotocol label switching (MPLS) is an IETF (Internet Engineering Task Force) specification currently under development for layer 3 switching. MPLS uses fixed-length labels to represent optimized traffic paths that can be quickly examined by a router. This approach is being developed to simplify packet forwarding, in order to lower the cost of high-speed transmission, improve forwarding performance, and better integrate QoS in the Internet. MPLS forwarding allows "aggregate forwarding" of user data, so that data can be forwarded as a unit in a stream that follows a single path. Although the initial MPLS effort is focused on IPv4 and IPv6, the core technology will be extendible to multiple network layer protocols.

5.2 Connecting Legacy Networks to New Network Structures

The Internet provides a way for businesses around the world to communicate with customers, partners, suppliers, and employees. New distributed business models are being introduced based on virtual workgroups that are connected logically rather than physically. TCP/IP has emerged as the unifying protocol for electronic business applications. Users connected to legacy networks with various legacy protocols must have access to each other through the TCP/IP-based Internet and private intranets. This requires support for end-to-end interworking through multiple protocols and media types. In addition, security becomes an issue when a closed user-group enterprise network is connected to the open Internet. A firewall may be needed to protect legacy networks from unauthorized access and tampering.

Installing a TCP/IP protocol stack in each workstation on the legacy network is one option for achieving interconnectivity with the Internet and its applications. Another approach is to use gateway internetworking software. The typical configuration (figure 5-2) consists of a TCP/IP link to a legacy protocol gateway server, with client software running on each workstation. The TCP/IP application (e.g., a Web browser) opens a connection with the client software, which then communicates with the Internet gateway server using the legacy network protocol.

Tunneling is often used so mobile users can access legacy corporate networks through the public Internet (figure 5-2). Tunneling is a technique for transporting data structured in one protocol format, within the format of another protocol. The legacy protocol is encapsulated within TCP/IP packets. Tunneling is used in ATM networks to maintain the functions of specialized legacy protocols (e.g., for encryption or authentication) while taking full advantage of native ATM services.

5.3 Making Legacy Applications Accessible in the New Environment

As new methods of data access are introduced, companies must maintain their older legacy applications and access methods for some period. This implies that data will be simultaneously accessed by both the legacy and new methods. For example, relational databases on a host system may be read and updated by legacy transaction systems. At the same time, a new electronic business application may be extracting data from those same databases and sending it down to a Web server to make it accessible to Web clients. These Web clients may be on the Internet or an internal enterprise intranet.

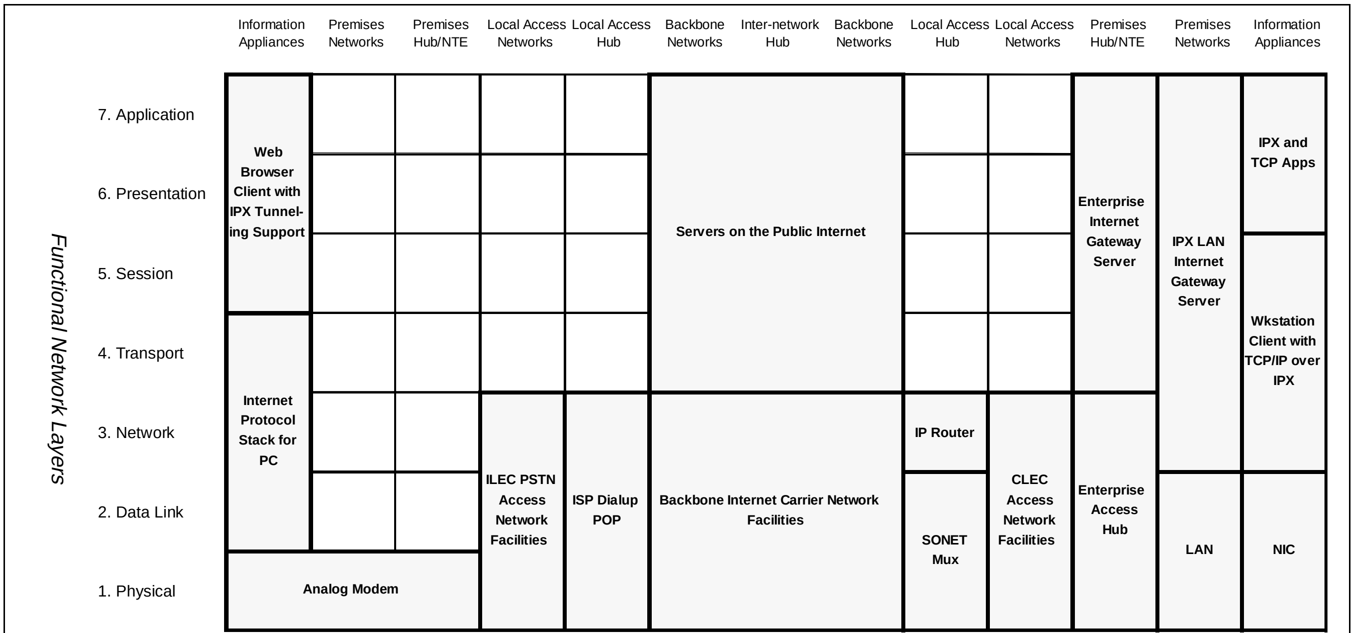

Figure 5-2 IP Tunneling and Web Access from IPX-Based LAN

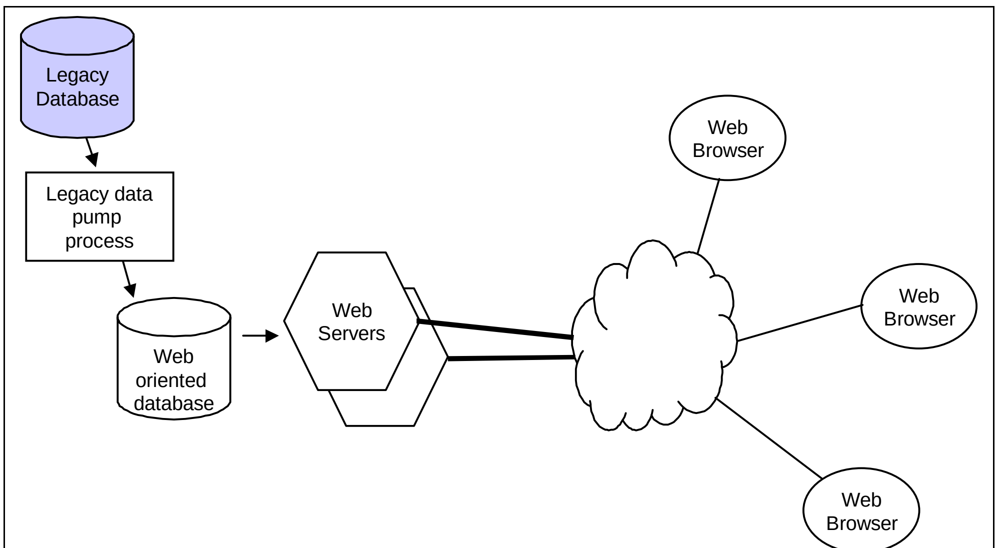

Figure 5-3: Internet Access to Legacy Data Systems

Software products are available to let Internet/intranet users access mainframe-based legacy applications through an Internet client such as a Web browser. These products allow the older, character-based screens to be replaced with graphical user interfaces. The typical configuration is to install a Web-enabling server that sits between the mainframe and the Internet/intranet. The server also provides access authentication, authorization, and other security features. The use of browsers equipped with Java virtual machines eliminates the need to install emulation software on the client workstations.

"Data pump" software lets Web servers access existing legacy relational database systems such as IBM DB2, Oracle, Sybase, Informix, and others in a controlled and managed way. In the typical configuration (figure 5-3), data are extracted and placed in a Web-accessible server. Data replication and synchronization must be managed to fit application needs (i.e., for static or volatile data).

The future NII/GII will continually adapt to satisfy the needs of the user community. Support for legacy protocols and applications will be a requirement, as will be the integration of new applications. Presently, natural competition is occurring between *ATM-centric* solutions for legacy protocols (such as MPOA) and *IP-centric* solutions (such as tunneling). The ultimate solution may not match any of the current ones exactly, but may use new technologies to incorporate the best attributes of all the existing proposals.

6.0 Voice over Packet-Switched Networks

Although the value of supporting all means of communications on an integrated services network has long been recognized, separate voice and data networks have been deployed due to the different characteristics of real-time speech and interactive data applications. The public switched telephone network (PSTN) employs circuit-switched technology to provide continuous point-to-point connections for real-time audio with a guaranteed quality of service. Modern data networks are based on connectionless store-and-forward packet technologies, predominantly IP, with little emphasis on QoS. Nevertheless, the potential economies of an integrated services environment are pulling the worlds of voice and data communications together as investment plans are made for the next generation of information infrastructure. Considerable research effort has identified packet technologies that can effectively support end-to-end delivery services for data with real-time characteristics, such as interactive audio and video. Today, several factors are driving the development and deployment of voice services over IP-based packet data networks.

6.1 Potential Cost Savings

Widespread availability of low-cost Internet access spawned early voice-on-the-'Net experiments as a cheap alternative to long distance telephone charges. Today, voice quality can still be unpredictable over the traditional best-effort public Internet, but the potential has been demonstrated. Voice over IP quality is more predictable in a private enterprise network environment where QoS levels can be guaranteed. Moreover, corporations have an incentive to use any excess capacity in their data networks to carry internal voice and fax traffic in order to eliminate recurring long distance charges from telephone carriers.

Regulatory anomalies for public telephone carriers have created an opportunity for the emergence of a new type of non-regulated telephone company which uses voice over IP technology to provide phone-to-phone voice services. In the United States, regulated long distance telephone carriers, Inter Exchange Carriers (IXCs), pay regulated local telephone companies, Local Exchange Carriers (LECs), an access charge of approximately three cents per minute on each end to reimburse the latter for terminating or originating a call. The Federal Communications Commission (FCC) has determined that enhanced service providers, including Internet service providers (ISPs), are not subject to pay either this access charge or the Universal Service surcharge. These exceptions enable ISPs to offer the flat-rate access services that have been instrumental in the growth of the Internet. New telephone carriers are emerging to take advantage of these same regulatory exceptions to offer discounted long distance services. The potential cost savings are even more pronounced for international telephone calls where current tariffs have little relationship to the cost of providing service.

6.2 Packet Voice Applications

The most basic IP packet voice application (figure 6-1) is to use the Internet to support real-time audio conversations between computer users (e.g., PC to PC). The underlying idea is to treat digitized voice as just another data type. All that is required for basic service is a multimedia PC with a microphone and speakers; compatible packet voice software to digitize, compress, and packetize the audio voice information at each end; and the IP address of the other party. By itself, this configuration has very limited capability. To make the service useful, there must be a way for the calling party to find out if the called party is available (on-line) and determine his or her IP address. These functions are provided by “address servers” which allow end-users to find one another. IP voice users register with the server, which provides a clearinghouse function for the broad user community. The address server is shown as deployed in the backbone network to emphasize that it provides a network-wide service — and requires appropriate network resources. This configuration, which provides an alternative voice network for a limited set of users, is the most basic form of IP telephony, but without the telephone.

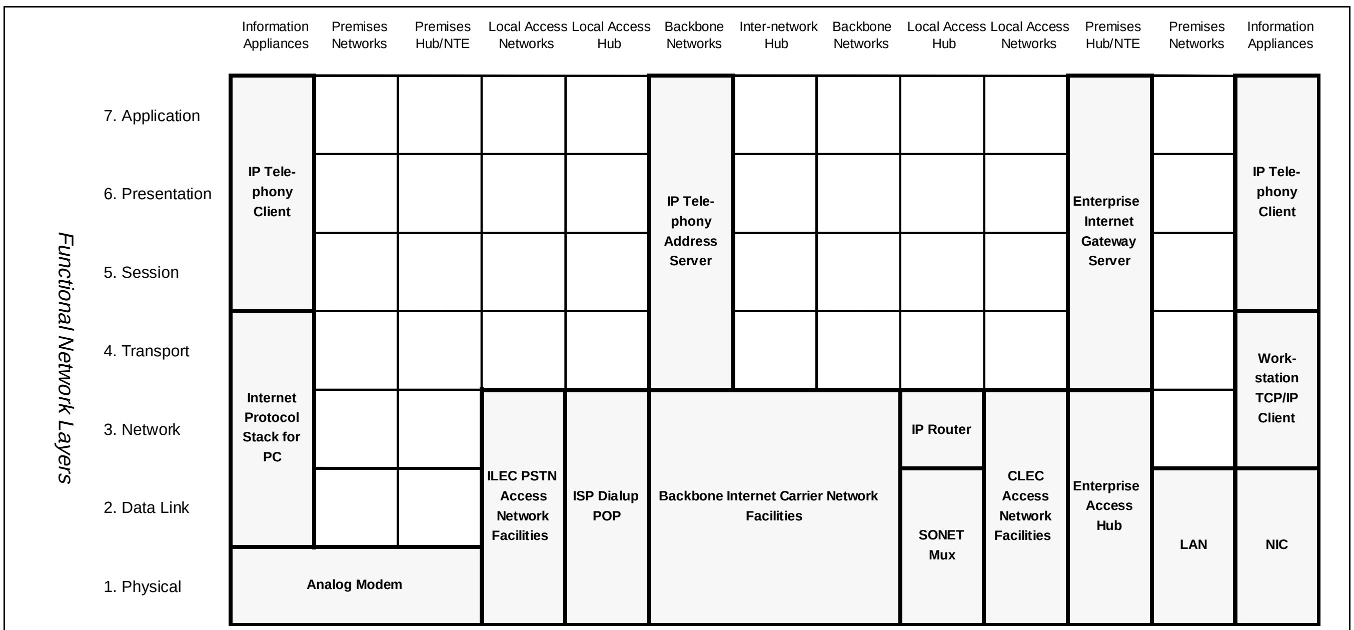

Figure 6-1: Basic PC-to-PC Internet Telephony

A next level of capability arises when a voice call can be initiated from a PC to a telephone on the PSTN. The PC-to-telephone application requires a gateway that connects the Internet to the PSTN (figure 6-2). This gateway can make a voice connection to any telephone on the PSTN as requested by the calling PC. The gateway takes a standard telephone analog voice signal, digitizes and packetizes it, and routes it to the destination computer over the Internet. This process is reversed (computer to telephone) to support two-way voice connections between a multimedia computer and a telephone. The gateway is shown at the local access hub to illustrate that the closer the gateway is to the terminating telephone, the larger the potential savings in long distance charges.

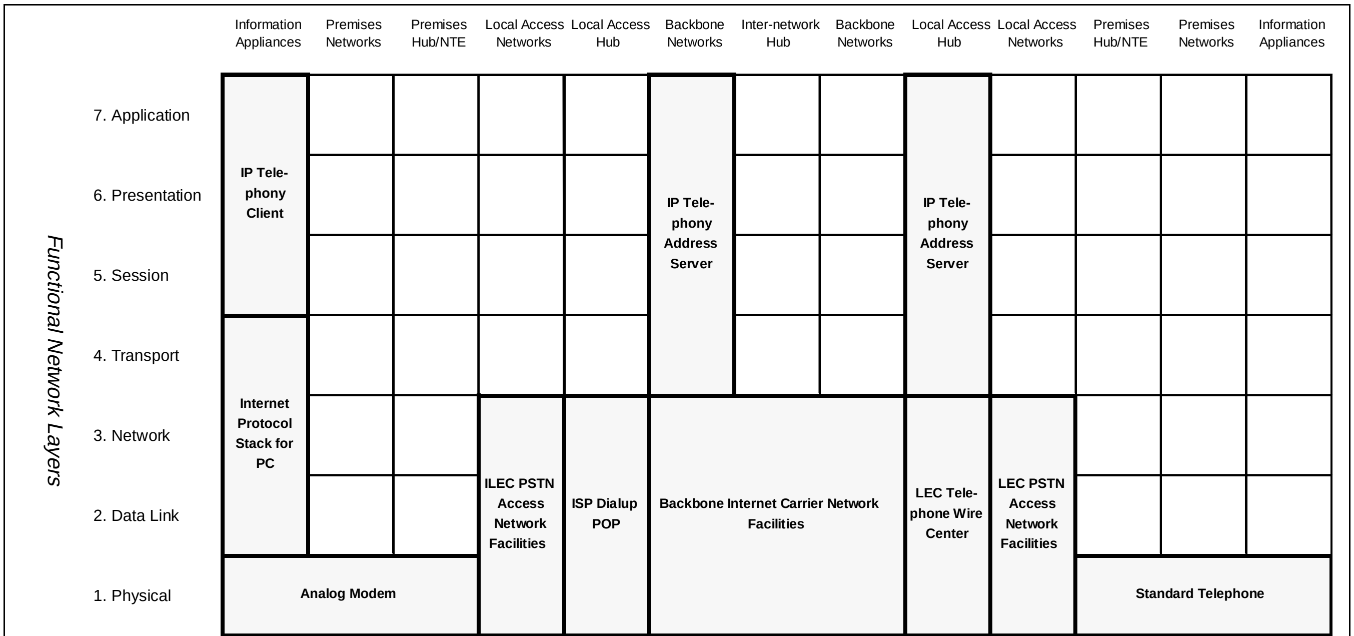

Figure 6-2: PSTN-to-Internet Telephony Gateway

The gateway is easily able to handle a voice call from a PC to a given telephone using the standardized worldwide telephone numbering plan, call control mechanisms, and embedded signaling and directory systems of the PSTN. However, to initiate a call from a telephone to any PC on the Internet is more complex because most Internet users do not have a fixed IP address. This is because most of these users dial in occasionally to an ISP, and an IP address is assigned to them on a temporary basis. To remain continually reachable, the end-user appliance will need to have a unique alias address equivalent to a telephone number. The IP network must be equipped with a supporting addressing and directory services infrastructure to map this alias telephone number to the user's current IP address. Efforts are under way to identify and standardize economically viable technical solutions to this problem. Until the problem is solved, telephone-to-PC calling will be limited to closed user-group applications, such as the integration of PBX and data network voice services at the enterprise level.

A single gateway allows computer-to-telephone connections. Adding a second gateway (figure 6-3) allows direct telephone-to-telephone connections using the Internet as an alternative to the PSTN IXC networks. Users make a simple telephone call to their local gateway and then dial through to the called party. In some situations, a telephone subscriber can designate an IP telephony carrier as their primary long distance company.

By locating IP telephony gateways in numerous local access hubs, the new IP telephony carriers can bypass the regulated telephone networks as much as possible and avoid paying regulated access charges to the local exchange carriers. Most carriers pass these cost savings along to their retail customers as lower per minute long distance prices.

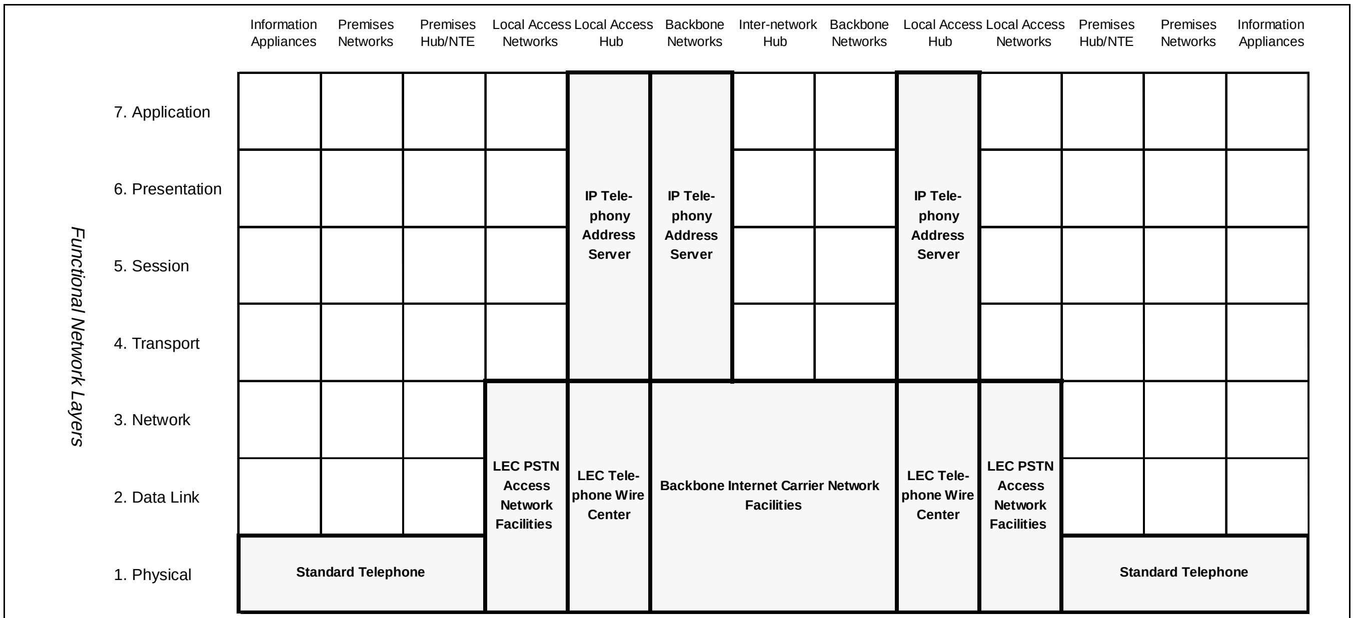

Figure 6-3: Telephone-to-Telephone Connections Using the Internet

IP telephony technology is penetrating enterprise networks driven by the potential for immediate cost savings. Operating separate voice and data networks is more expensive than operating a single multiservice network. There is potential for additional savings from reduced long distance charges and more focused capital investment strategies. IP telephony functions are being integrated into WAN access equipment to support data and voice integration in enterprise networks. PBX suppliers are also developing IP voice ports and features. A common application (figure 6-4) is to use excess capacity on a corporate data network to support internal tie lines between PBXs.

6.3 Meeting Service Challenges with Packet Voice Technology

There are a number of challenges that must be met in order to advance the market for Internet telephony. Voice *quality* is the most critical issue for service acceptance. IP telephony voice quality has improved significantly due to improvements in voice coding and lost packet reconstruction. Latency (delay in packet arrival) is another factor in the quality of voice conversations. Current IP telephony implementations may have delays exceeding 250 msec, which makes them sound much like calls over satellite links. Improvements in gateway designs will reduce the delay, but the primary cause of delay is the nature of best-effort packet network technology used in the Internet. The increased delay places additional requirements on the IP telephony gateways to perform echo cancellation. Service providers can reduce and control the delay between gateways through the use of private links and bandwidth utilization control measures. Techniques are also being developed to provide better real-time performance over packet networks, but it will take time to upgrade the public Internet. ATM's QoS capabilities make this technology quite attractive in networks with significant real-time traffic. Several alternative techniques are available for IP working over ATM networks.

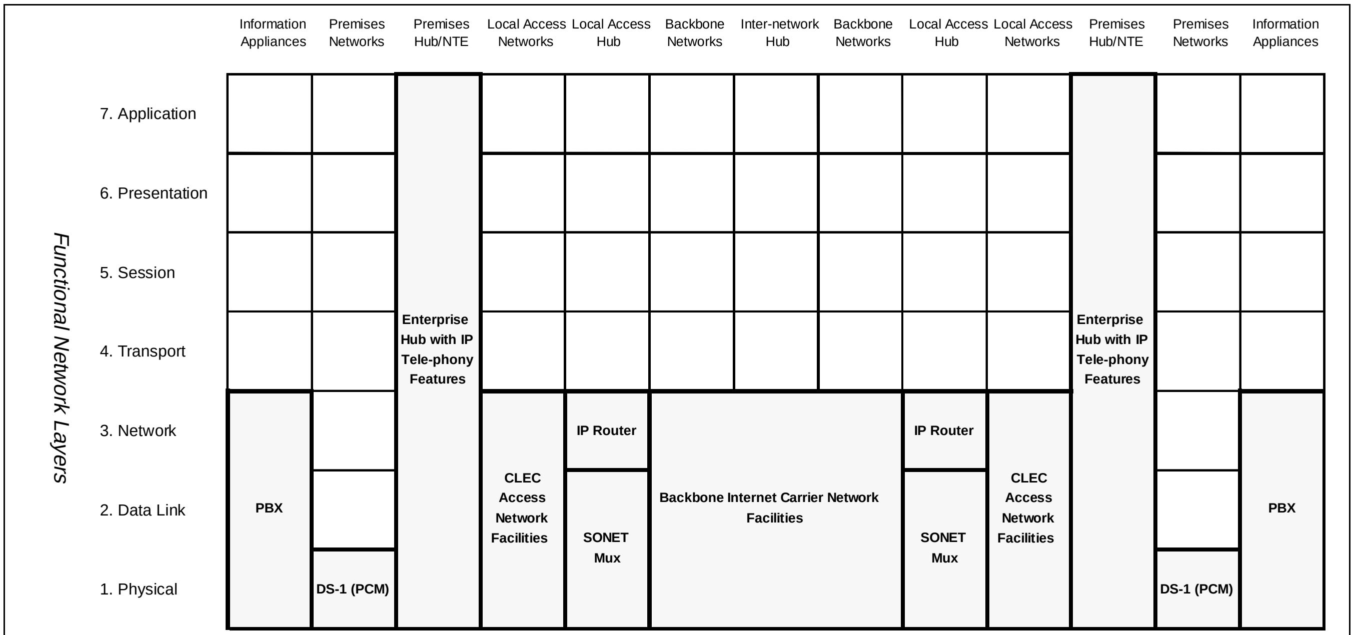

Figure 6-4: PBX Interconnection with IP Telephony

Service *transparency* for embedded voice network services and applications based on touch-tone signals is another issue for IP telephony. Coding and packetization distorts dual-tone multifrequency (DTMF) digits and makes them unrecognizable at the far end. IP telephony gateways must detect DTMF digits locally, suppress their transmission in the IP links, and regenerate them on the remote telephony side with very little added delay.

Directory services require improvement. In the PSTN, if a person's telephone number (PSTN address) is known, a call can be initiated. On the Internet, a user's IP address typically changes each time that user connects to the Internet. A directory service that can keep up with the relationship of user name and current IP address is an unresolved issue. A related challenge is how well the directory and routing services will scale as the number of users grows and still support the short dial-to-connection times associated with the PSTN.

Carrier grade service standards must be established. The ultimate commercial viability of IP telephony depends on the ability of the technology providers to provide products that can support carrier grade services deployable on a small to large scale. These products must support all aspects of the service business including service provisioning and management, service fulfillment, user authentication, security, directory services, routing control, system integrity, billing, revenue collection, and accounting.

6.4 Standards and Interoperability

IP telephony systems have limited value as isolated islands. Their highest value is from interworking with each other and the traditional PSTN. Work is under way in the industry to develop standards for IP telephony. H.323 is the ITU specification that defines packet standards for terminals, equipment, and services for multimedia communications over LANs. It has been adopted as the basis for standards for multimedia communications over any packet network, including the Internet. The H.323 standard specification includes the terminal client software, the gateway functions, a conferencing bridge for audio and video, and the connection manager. H.323 defines several voice coding-decoding (codec) standards including GSM, G.729, and G.723.1. IP telephony systems are expected to be capable of supporting multiple voice coders as market choices and technology evolve. Version 2 of H.323 focuses more on the requirements of IP telephony; and work on version 3 is under way. The standardization effort for the higher level functions required to achieve carrier grade service is just getting started, and much work remains.

Full interconnection and services interworking with the PSTN is essential for widespread acceptance of packet voice service as an alternative to the traditional telephone network. This will require standardization efforts to harmonize addressing structures, call control mechanisms, and directory services. It will also be necessary to reach common agreements about how to port the many optional calling services available on the PSTN to an integrated environment.

6.5 Future Directions

IP networking provides a way to combine the traditional telephone environment with the new world of interactive data networking and to move closer to the vision of integrated communications services and applications. Today, most packetized voice development attention is focused on voice over IP technology. Critical issues concerning service performance and network scalability remain. We all depend on and take for granted the high reliability of our telephones and telephone network services. Achieving similar reliability results with data terminals and data networks will be difficult and costly. Practical QoS approaches for supporting real-time voice over the public Internet must be in place before this technology can scale to the level of the PSTN. One best-effort service class appears to be insufficient to support a mixture of demanding multimedia applications. In the near term, IP telephony technology will see growing use in closed networks where performance can be more readily managed.

The regulatory anomalies and rate arbitrage that stimulated the early interest in voice over IP and IP telephony are not a sound basis for making long-term investment decisions. Rates and regulations are subject to revision at any time. The more fundamental driver for continued development of packetized voice technology is the need for voice, data, and video to coexist in the next generation of networked applications, including multimedia and multiparty communications. The ultimate value of the new technologies will be through applications that provide new features and functionality. This may be adding voice conversation capabilities to Web pages and electronic commerce applications, or universal messaging systems that combine voice mail, fax, e-mail, and video messaging. Universal messaging gateways, for example (figure 6-5), present a new service market opportunity for communications carriers and information service providers

In the enterprise, multiservice networking that includes voice services will enable more effective strategic business applications convergence in such areas as enhanced call center operations, internal help desk support systems, and networked collaboration environments. As multimedia applications grow in importance, the enterprise LAN and PBX will begin to merge, bringing IP telephone sets to many desktops.

In the residential market, traditional POTS will continue to provide the primary telephone line well into the future. With the availability of high-speed access technologies, however, there will be a growing market for secondary voice lines and services using IP telephony technologies.

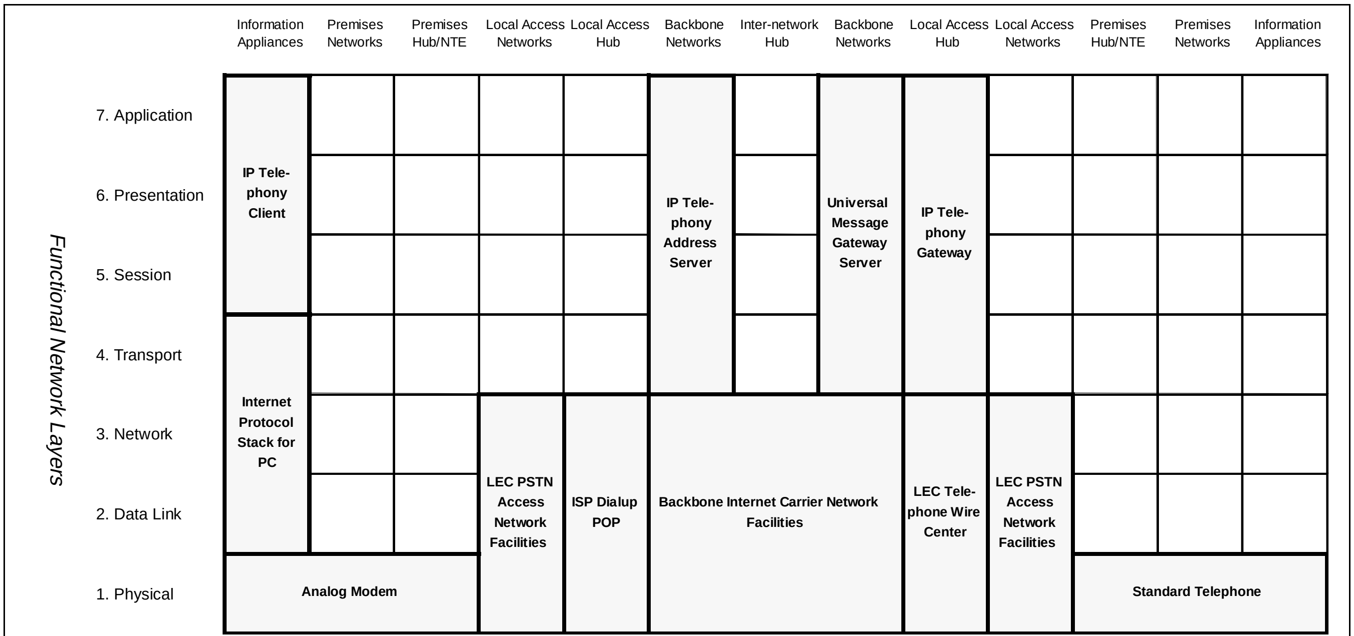

Figure 6-5: Network-Based Universal Messaging Services

7.0 Optical Transport Technology

Demand for information transport capacity continues to increase. To respond to this increase, carriers have installed fiber optic cable and have begun to deploy dense wavelength division multiplexing (DWDM) systems. In DWDM systems, optical fiber carries light of more than one color at a time. Since light signals do not interfere with each other, they represent independent information-carrying channels.

7.1 DWDM Networks

DWDM systems create an additional layer in today's telecommunications networks. This layer has some of the properties of fiber optic cable in that it uses light to carry information from point-to-point. But this layer has additional properties as well. For one thing, unlike the physical fiber, a channel can be switched. This switching takes place when light of one color is separated from light of other colors within a fiber and is redirected elsewhere. To accomplish this function at the fiber level, the carrier would have to break and replace the fiber, or arrange for a switching function to accomplish the same effect.

The second major difference is that each individual channel has less bandwidth than the fiber as a whole. This means that the carrier can “groom” the fiber bandwidth of the network for maximum efficiency. Once grooming is introduced, the carrier can begin to consider using available colors for protection and other purposes. In principle, it is possible to accomplish many of the network functions dealing with reliability and protection at the optical layer, rather than at the data link layer (e.g., SONET).

Since optical networks are format and rate independent, they play a special role in the transition from networks designed for voice services and circuit-switched protocols to packet data networks. The use of an optical layer can facilitate the use of multiple protocols over the same physical facility. Such flexibility gives carriers the ability to mix and match various protocols and systems.

DWDM will clearly advance in point-to-point networks as a means of increasing capacity. The continued need to support legacy networks will extend an advantage to DWDM networks. This is especially true if LAN technology is extended into the metropolitan area network (MAN) environment, and older protocols must be supported.

7.2 Skipping Layers

Since DWDM introduces a new layer, the prospect of eliminating other layers in the network where there are duplicating functions tempts system designers. However, various layer-skipping possibilities are limited by carrier strategy and prior equipment investment.

In the case shown in Table 7-1, data services have separate dedicated capacity, as does voice. The networks share a common pipe, but there is no pooling of capacity requirements. As this is relatively inefficient for low levels of demand for data capacity, it leads to a first round of capacity problems.

	IP (data)	SS7 (voice)	SONET	DWDM pp	DWDM networks
Application	✔				
Connection	✔	✔			
Grooming			✔		
Protection			✔		
Multiplexing			✔		
Transport			✔		

Table 7-1: Legacy Solution for a Mixed Services Network Carrying Data

In the case shown in Table 7-2, the capacity requirements for voice and data are pooled using ATM as above, but DWDM is added to increase transport capacity. Voice and data are integrated at the connection level. In other circumstances, IP augmented with multiprotocol label switching and RSVP may serve as the integrating layer. In either case, the role of DWDM networks beyond providing more channels is indeterminate.

	IP (improved)	ATM	SONET	DWDM pp	DWDM networks
Application	✔				
Connection		✔			
Provisioning			✔		
Protection			✔		
Multiplexing			✔		
Transport				✔	

Table 7-2: Possible Solution for Mixed Service Network

7.3 Interaction with Other Protocols and Standards

Since the optical layer is not "aware" of the specific protocols carried at higher layers in the network, a carrier emphasizing optical layers could be considered analogous to a donor of type O+ blood. The optical layer is the universal donor; any protocol that can be converted into intensity modulated[1] light can propagate in the fiber with limited interference with other light signals.

7.4 Transition to Optical Networking

The transition from electronic networking to optical networking appears likely to occur in stages as shown in Table 7-3. Non-value-added functions in the network (those that do not add value from the customer's point of view) will gradually give way to photonic equivalents. This process is under way.

	Electronic function	Photonic function
Phase 1 – Transport functions within network replaced	Regeneration	Optical amplification
Phase 2 – Transport functions at edge of network replaced	Digital multiplexing terminals	Wavelength multiplexing terminals
Phase 3 – Network functions within network replaced	Switching and cross-connect	Optical cross-connect, Add/drop multiplexing
Phase 4 – Network functions at edge of network replaced	Terminal electronics, digital switched, network control	Wavelength selective addressing

Table 7-3: Transition to Optical Networking

[1] Intensity modulated light means that information transport occurs when light intensity or brightness changes in correspondence to the underlying data. In simplest terms, a bit (or mark) is an optical pulse; a space (or mark) is a dark time slot. Note that there are data-carrying formats that use continuously varying light intensity modulation.

8.0 Middleware Technology

Middleware functions are the "glue" that binds client and server applications. Middleware products provide essential functions for building and operating large-scale networked applications and information systems in a heterogeneous environment. In a networking context, middleware is the software that helps applications obtain services from service-providing entities in or connected to the network. Winsock, for example, is middleware with which applications on Windows-based computers use communications services such as TCP/IP without having to “know” the details of those protocols.

A generic approach to middleware, becoming increasingly important to networks and overall information infrastructure, is the distributed object system. The main examples today are the Common Object Request Broker Architecture (CORBA) standardized by the Object Management Group, and Microsoft’s Distributed Common Object Model (DCOM), together with their respective libraries of object services. Distributed object system middleware is a software platform for communication between applications and service-providing objects that hides the details of using system services across multiple computers, operating systems, and communications protocols (figure 8-1). This middleware supports applications such as:

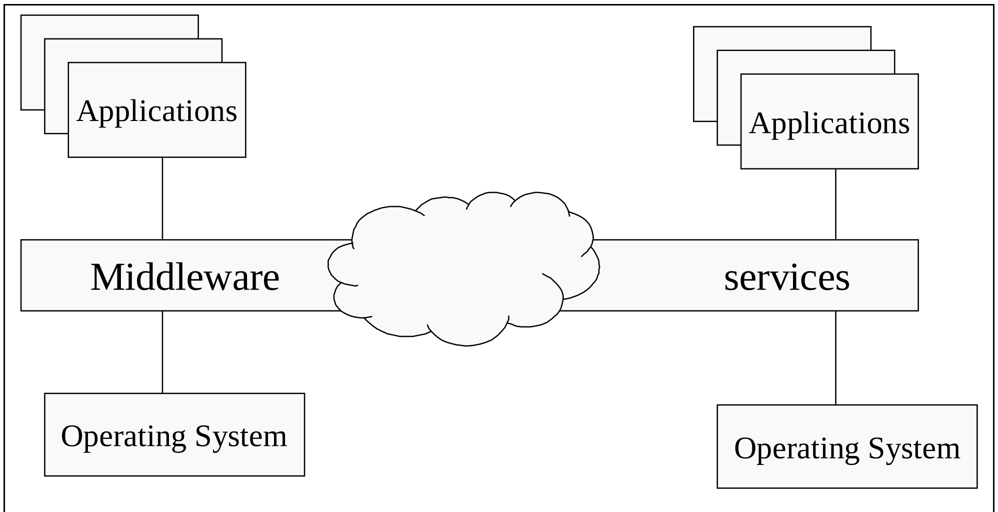

Figure 8-1: Middleware Fits between Application and Operating System

- the creation of QoS-sensitive *virtual services*, for example, a customized multiparty, multimedia communications session, as objects invoked through Web browsers;
- secure and user-friendly access to a diverse set of financial services as a single integrated environment, despite the services being provided on different computing systems and written in different computer languages; and
- management, through a single Web-based interface, of network elements and systems using diverse legacy protocols.

Distributed object middleware services include communications, naming and directory services, object life cycle management, security and authentication, notification service, and various development and debugging tools. These services are needed in applications such as database management, transactions processing, resource management, fault recovery, and the support of multiple communications or business protocols. Software reuse is implicit in these services, since they do not have to be reprogrammed as new applications are written. Building applications on a middleware infrastructure has numerous advantages for the application developer including increased programming productivity, lower development risks, and more portable applications.

8.1 Middleware Services Today

We have described one category of middleware services, based on a communications mechanism derived from the remote procedure call, that is particularly important for applications in or making substantial use of networking. There are, however, different forms of middleware, which address various application problems. Middleware products tend to fall into one of five categories based on the mechanism used to pass information between application processes.

Remote procedure calls (RPCs) allow a program (client application) to invoke a procedure that is located remotely as if it were local, i.e., in the same machine. The client waits until a result is returned from the server before using RPC again. An interface definition language (IDL) is used to define the service and is compiled into the languages used in the server and client applications. RPC is a simple mechanism that leaves most system-level details to the application developer. For this reason, RPC by itself is not attractive for complex applications.

Distributed object systems, such as CORBA and DCOM, can be thought of as object-oriented RPCs with added capabilities such as the object services alluded to above. With object-oriented environments becoming predominant for network management and control, as well as many end applications, distributed object systems provide useful platforms upon which a wide range of services architectures can be built. In general, both CORBA and DCOM run on top of TCP/IP (figure 8-2), and so are based on a reliable underlying communications environment.

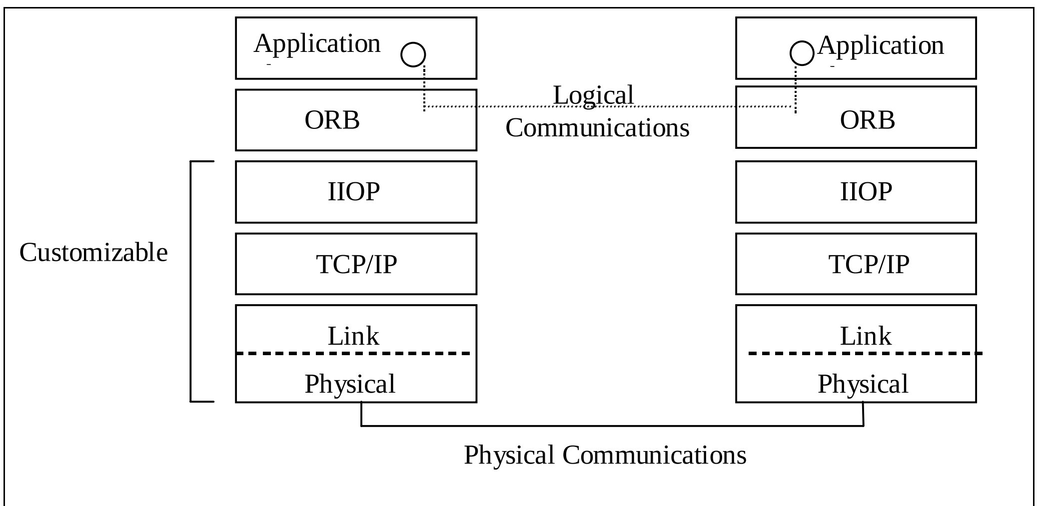

Figure 8-2: CORBA Middleware Running on Top of a TCP/IP Protocol Stack

Concerns have been expressed about the performance (specifically, the computing response time) of distributed object systems. The problems which led to these concerns appear to have been in implementation rather than fundamental concepts. Distributed object systems are being developed for demanding real-time applications such as communications network control. Scalability to tens of millions of objects is also a concern because of directory problems and the difficulty of setting up reliable communications between any client-server pair. This scalability issue is in principle similar to providing universal Internet service, so it is believed that it will be addressed as the information infrastructure evolves.

The Java virtual machine offers a similar RPC-based communications middleware. Java and CORBA provide platform (operating system) independence. Java-based distributed object systems require all applications to be written in Java. The CORBA distributed operating system allows applications to be written in any language, including Java, for which a CORBA communications mechanism — the object request broker (ORB) — is available.

Message-oriented middleware provides mechanisms to pass information between processes (programs) via messages rather than interactive communications. Messages are sent asynchronously so the sender does not have to wait for a reply. One popular message-oriented middleware mechanism is **message queuing**, whereby programs communicate using a queuing service. The programs put outgoing messages in a send queue and get incoming messages from a receive queue. Because messages are protected while in the queue, message queuing is very good for applications involving mobile users who are not always connected to the transport network. Another popular messaging mechanism is called **message passing** or **publish-subscribe**. Programs subscribe to subjects of interest and receive any messages sent (published) on those subjects. Publish-subscribe is a highly scalable architecture since it can work with message distribution mechanisms that do not require the message sender to have a direct connection to all of the destination programs. The mechanism is thus useful for real-time replicated database updates. Most commercial products using publish-subscribe assume that the underlying transport network is not reliable, and therefore include security, error recovery, and message priority services.

Distributed transaction-processing monitors track a transaction as it passes from one stage in a process to another to ensure that the transaction is completed or the appropriate action is taken in the event of an error. Historically, these monitors have been reserved for transactions such as banking and sales, that require the properties of atomicity, consistency, isolation, and durability.

Database access middleware integrates data residing in multiple databases throughout an enterprise. A program makes a request, in structured query language, to a database gateway supplied by the middleware vendor. The gateway then makes a direct request to the target database in its native dialect. The database middleware architecture may not be suited to high-performance applications if the gateway becomes a bottleneck with many programs vying for access. Distributed object systems can also provide translation services for database access, exemplified by GDMO/CORBA IDL translation of management information between ISO common management information protocol network management systems and CORBA-based management systems.

Commercial middleware products first emerged when it became clear that a "build-it-yourself" development approach could not keep pace with the increasing complexity of large distributed applications operating in heterogeneous environments. Despite the improvements in commercial middleware, business systems are still dominated by proprietary implementations of point solutions to specific application problems. Commercial middleware components are not well integrated with each other, even for products from the same vendor, although this situation is rapidly improving. The growth of business and commerce over private intranets and the public Internet is underscoring the need for reliable, interoperable middleware services and solutions based on open interface standards.

8.2 Expanding Role of Middleware

Over time, a commercial middleware market has developed to support the explosive growth in enterprise-wide distributed business environments. Business users today expect to have instant access to information from anywhere in the enterprise. With the Internet, this is becoming an expectation of consumers as well. This increasing reliance on networking is exposing the vulnerability of ad hoc approaches for supporting anticipated networking growth and complexity with robust and reliable systems.

Meanwhile, information technologies continue to advance at a rapid pace, raising a wealth of new possibilities and user expectations. Ever-increasing computing power means that applications will become more complex and comprehensive. Higher bandwidth at lower cost will accelerate the move to decentralized and distributed organizational structures. Distributed and mobile staff must be connected to the enterprise network and to each other to access information and applications and to share tasks. Cross-enterprise collaboration and interaction with customers are also distributed networked applications.

The interoperability of applications written in different languages and running in a variety of computing environments (operating systems) will be supported by middleware. For example, if the distributed environment does not have to accommodate legacy applications in different languages, Java applications can run on virtual machines, realizing at least operating system independence. Object services such as naming, object life cycle management, security, and notification will become a valuable set of reusable components that will speed and simplify the writing of applications. Additional object facilities, also reusable software, can be written for particular application areas. Still, some very practical problems result from rapid technology change. Very few individual enterprises have the skills and/or resources necessary to keep up with the changing technology by themselves. New approaches are needed to take full advantage of the opportunities these changes present.

Many believe that middleware concepts are key to a new system metaphor that will allow us to maintain system reliability and performance through periods of rapid change. In this view, an expanded middleware infrastructure will take over functions from operating systems and applications programs, allowing users to concentrate on what brings value to their applications. For example, in an application where information is gathered, stored, organized, filtered, analyzed, and presented, the application programmer will concentrate on information filtering and analysis, leaving the other functions to the middleware infrastructure. Middleware will link diverse digital information types to form an integrated view based on the needs of the individual user, and present this view in a form appropriate to the information appliance that individual is currently using. Rather than own all of their computing and storage resources, users will lease or rent, from a service provider, virtual system resources — represented by object abstractions in a distributed object system. These resources will be provided as value-added services of communications networks, as a public computing utility, or as specialty “branded” services. The middleware infrastructure will help users identify their needs, locate resources, and negotiate terms and conditions for use.

We can illustrate alternative (or parallel) scenarios of how this expanded middleware infrastructure might be deployed by examining two networking extremes. At one extreme, a “dumb” communications network is a ubiquitous connector carrying a mix of large and small information flows among users. Customers at the end points, for example, using smart cards together with computers running Java virtual machines and applications, connect through a ubiquitous network to check their account status. Middleware, maintained by the companies with which the customers have their accounts, provides the graphical user interface, search tools, and security and authentication mechanisms. The system all works independently of the physical accessing devices. The business is able to serve a completely unmanaged and heterogeneous base of customers.

At the other extreme, the “intelligent” network itself is the service destination. The network has middleware-enabled embedded applications and content, for example teamware and group collaboration environments. Using these middleware-enabled network services, a group of small companies can join to collaborate on projects, using tools and resources that collectively are competitive with those available to the largest enterprise. The underlying infrastructure technology is essentially the same at both extremes. The real difference is the business model selected and who owns, deploys and maintains the middleware.

Middleware infrastructure components will be deployed throughout the distributed network of networks that is the foundation of the NII, as shown in figure 8-3. In this example, the future wireless terminal (A), equipped with mobility middleware and possibly a virtual machine, becomes a multipurpose portable information and communications center. It can be dynamically programmed to be a telephone, a pager, a messaging center, a portable ATM that can load your smart card with money, or a portable PC, depending on the immediate application needs. Similarly, enterprise workstations equipped with client middleware (B) will invoke and control applications both on and off the enterprise network, leaving staff to concentrate on customer service and partner relationships. Enterprise middleware services (C) will enable authorized users to access corporate information without having to have preloaded display, security, and other necessary software. Network middleware services (D) will enable users to find each other, support advanced billing systems, allow highly personalized communications service packaging, and open numerous opportunities for innovative network services. This future middleware infrastructure can be the means to satisfy the need for scalability, applications interoperation, and operations outsourcing.

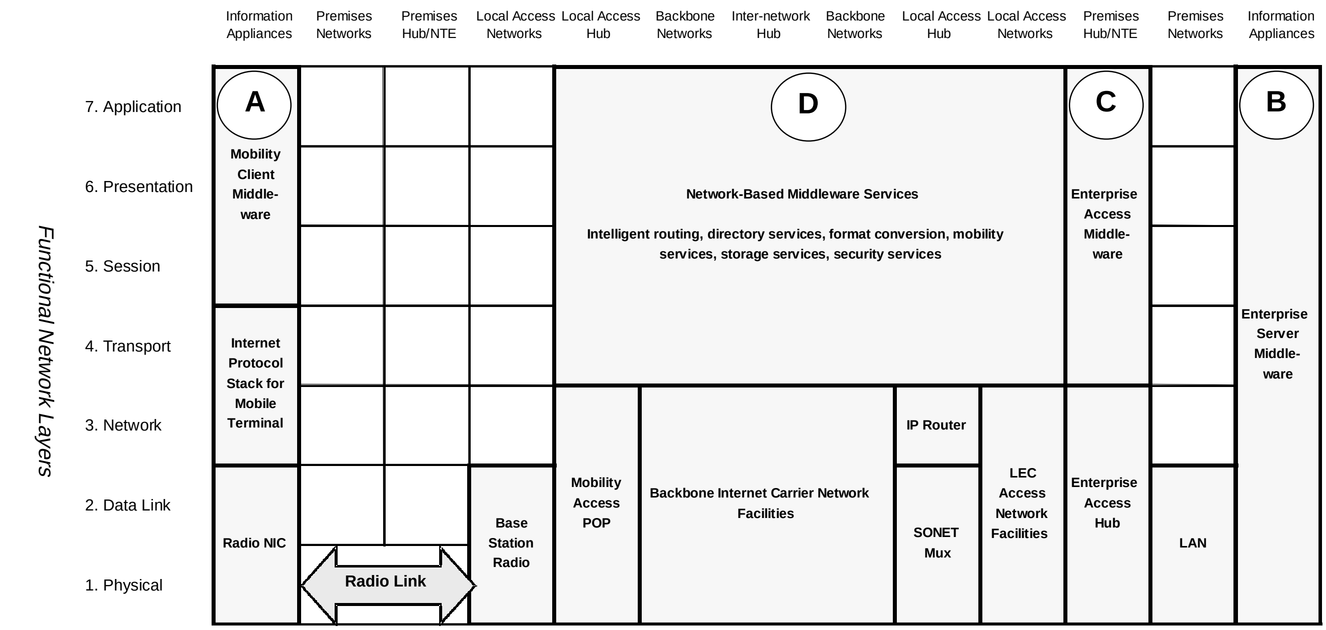

Figure 8-3: Middleware Has Application throughout the Distributed NII

9.0 Status and Outlook

We conclude this examination of current and emerging networking technologies by returning to the guiding principles and goals for the new NII/GII first outlined in the XIWT white paper *An Architectural Framework for the National Information Infrastructure* and assessing progress towards those goals.

9.1 Widely Available to All Americans

Widespread and affordable access to the public Internet has demonstrated to many Americans the potential of networked applications of information technology. Competition in the computing arena has fostered innovation. Prices for appliances (both personal computers and specialized networked appliances) are falling, and larger proportions of the world’s population are gaining access to them. However, functional and performance limits of the available infrastructure have limited actual use to simple applications such as e-mail and Web browsing. DSL and cable modem technologies promise high-speed access services at affordable prices. Network carriers are steadily entering the marketplace, and competition is generating major investment and deployment commitments.

9.2 Operate in an Open Competitive Environment

Telecommunications regulatory reform is under way around the world. Markets are being opened to competition. While the early results are chaotic, technology advances are providing many opportunities for new players to enter previously closed markets. In many instances, new technology is being used to offer lower cost substitutes for existing services where market demand is known.

Technologies to support new integrated services and applications are emerging. New communications carriers are being formed to employ these new technologies in innovative multiservice business models. The underlying result of deregulation is market-driven movement toward integrated network services. During regulatory transition, embedded carriers and new entrants often operate under different rules. This is creating new, and perhaps short-term, market opportunities (e.g., discount IP telephony services). Nevertheless, competition is driving investment and innovation, and progress is being fostered.

9.3 Protect the Rights of Users and Stakeholders

Stakeholder rights are at the heart of current public policy debates about privacy, socially acceptable behavior, and accountability in the new networked environment. There is considerable tension in various public arenas around the world regarding the rights of users vis-à-vis society, as represented by the state. Issues include the balance of individual privacy and legitimate law enforcement needs; free speech versus protection from libel; and accessibility to information deemed harmful, such as pornography or political dissent.

Technology innovation is providing tools to protect stakeholder rights. Software technology can provide many infrastructure functions, at the middleware or application level, essential to protecting these rights. For example, middleware infrastructure can identify and authenticate users, support secure billing services, and manage intellectual property. End-user applications can provide filters and other information management capabilities. The question of who will control information and the technology remains open in most areas of the world.

9.4 Promote Interoperability and Open Standards

Open voluntary standards are recognized as the primary tool for building an interoperable scalable infrastructure. Progress is being made in accelerating standards-making processes to better match the pace of technology innovation.

Networked applications involve numerous technologies that must operate as an end-to-end system. Basic transport standards are now in place, but progress has been limited in achieving interoperability in the higher layers — for example, security services interoperability and directory services infrastructure. More active cross-industry cooperation is needed to identify and address these higher layer interoperability issues.

As always, vigilance is needed to prevent market distortions due to monopoly power and unfair business practices.

9.5 Provide Dependable High-Quality Services

There are today more service choices supporting more applications, and considerable progress has been made toward realizing the NII/GII vision. Individuals and businesses are routinely finding new and important uses for the Internet. Substantial investments are being made for new commercial services, and many firms are finding successful applications. However, service quality is occasionally problematic due to performance and reliability variations. Users of data networks cannot purchase a specific and guaranteed end-to-end quality of service over a generally affordable public data network. This has slowed the adoption of public data networks for important applications and has required expensive custom configurations. Moreover, growing system complexity raises serious issues of network management, as well as software safety and information assurance. Progress has been considerable, and the trajectory positive, driven by competition and its fruits, especially innovation and investment in network capacity. As societies and economies grow more dependent on the emerging infrastructure, this area will require significant attention.

9.6 Provide an Information Marketplace

An unparalleled number of new content providers sell information and provide services on the Internet. Measures such as the number of domains, and the ubiquity of domain names in advertisements and on business cards, indicate a steady growth of Internet users. So rapid is this growth that it is difficult to observe the effects of purported infrastructure gaps in security, intellectual property protection, and access reliability that are said to be slowing the commercial exploitation of new network applications. Parts of the problem involve finding workable business models for new media and content types. Nevertheless, this does not appear to be slowing either financial investment or public fascination with new content services. Emerging middleware tools that address upper layer functions appear to be keeping entry barriers low for smaller commercial providers and individuals. It is widely believed that a process of integrating existing applications to create new services is beginning, and will be the economic engine for future deployments and infrastructure investments.

9.7 Conclusion

The technological and business issues that must be resolved to enable the NII/GII vision will continue to require the best creative efforts of industry and the broader technical community. Cooperation in the creation of standards, and interactions among competing firms will be needed to sustain experimentation, innovation, and the successful operation of a highly distributed, continually scaling network of interworking networks and applications. Significant new R&D investments will be needed for research and, just as importantly, for the development of human resources to carry the vision forward.

This represents a challenge for which success is not yet guaranteed. Public policies can play a critical role in this regard: they can make or break the NII through their potential impact on interworking. In this light, XIWT urges the creation of *effective* public policy to sustain the progress of the past decade. This public policy should meet the following criteria:

- Policies should be developed in line with a vision of what the telecommunications, computing, and information industries should look like in 10 to 20 years.
- Policies should be consistently applied to all within regulated industries. Policy that is uneven, or that creates the opportunity for arbitrage based solely on regulatory irregularities, is flawed.
- For the telecommunications industry, which has traditionally been regulated, policymakers' vision of the future NII should be a yardstick for measuring the effectiveness of regulators, such as the Federal Communications Commission and its counterparts.
- Traditionally unregulated industries, such as content providers and computing hardware and software manufacturers, should remain unregulated. They should not be burdened with disincentives to innovation and growth so long as the marketplace operates in acceptable ways. These segments of the information industry are engines of social progress and wealth generation.
- Great care needs to be taken to enable innovation and not to burden technologies that serve broader purposes with unworkable requirements to solve perceived social problems.
- Policy in this era of unparalleled innovation, investment, risk taking, and experimentation must be flexible enough to accommodate current and continually emerging technologies and uses.

Public policy for the "information millennia" needs to be focused on the long-term principles and goals outlined above. There is little doubt that the "genie is out of the bottle," and the information infrastructure described in this paper will continue to develop apace — with or without effective public policy. However, the nature and utility of the infrastructure, and the benefit that society can derive from it, will be profoundly impacted, for good or bad, by the public policies put in place over the next few years.

10.0 Glossary of Acronyms

ADSL	Asymmetric Digital Subscriber Line
ATM	Asynchronous Transfer Mode
CLEC	Competitive Local Exchange Carrier
CORBA	Common Object Request Broker Architecture
CPE	Customer Premises Equipment
CSU	Customer Service Unit
DCOM	Distributed Common Object Model
DECnet	Digital Equipment Corporation Network
DSL	Digital Subscriber Line
DTMF	Dual-tone Multifrequency
DWDM	Dense Wave Division Multiplexing
FCC	Federal Communications Commission
GDMO	Guidelines for Definition of Managed Objects
GII	Global Information Infrastructure
GSM	Global System for Mobile Communications
HFC	Hybrid Fiber Coax
ICW	Internet Call Waiting
IDL	Interface Description Language
IETF	Internet Engineering Task Force
ILEC	Incumbent Local Exchange Carrier
IP	Internet Protocol
IPX	Internetwork Packet Exchange
ISO	International Organization for Standards
ISP	Internet Service Provider
ITU	International Telecommunications Union
IXC	Inter-Exchange Carrier
Kbps	Kilobits per second
LAN	Local Area Network
LATA	Local Access Transport Area
LEC	Local Exchange Carrier
LMDS	Local Multipoint Distribution Service
Mbps	Megabits per second
MMDS	Multichannel Multipoint Distribution System
MPLS	Multiprotocol Label Switching
MPOA	Multiprotocol over ATM
NII	National Information Infrastructure
NTE	Network Terminating Equipment
ONU	Optical Network Unit
ORB	Object Request Broker
OSI	Open System Interconnection
PBX	Private Branch Exchange
PC	Personal Computer
POP	Point of Presence
POTS	Plain Old Telephone Service
PSTN	Public Switched Telephone Network
QoS	Quality of Service
RPC	Remote Procedure Call
SNA	Systems Network Architecture
SONET	Synchronous Optical Network
TCP	Transmission Control Protocol
TV	Television
WAN	Wide Area Network
XIWT	Cross-Industry Working Team