Technologies and Systems for Access and Transport Networks

For a listing of recent titles in the *Artech House Mobile Communications Series*, turn to the back of this book.

Technologies and Systems for Access and Transport Networks

Jan A. Audestad

ARTECH HOUSE
BOSTON | LONDON
artechhouse.com

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the U.S. Library of Congress.

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library.

ISBN-13: 978-1-59693-299-9

Cover design by Igor Valdman

685 Canton Street
Norwood, MA 02062

10 9 8 7 6 5 4 3 2 1

To my wife Synnøve

Contents

CHAPTER 6

CHAPTER 7

Preface

Telecommunications has undergone a huge evolution during the last decade. The evolution has taken place in the political, commercial, and technological arenas at the same time.

In the political arena, the most important change took place in Europe in 1998, when all telecommunications was opened for free competition. Mobile communication had already been commercialized. Before 1998, telecommunications had been monopoly business, where the monopolies (or cartels) were operating alone in given geographical areas (e.g., a country or a city). Moving from monopoly to competition required strong market regulation by the government to prevent the incumbent from utilizing the market power it had built up on public investments for more than a century.

The liberalization of the market had, of course, commercial implications. However, building traditional telecommunications systems required huge investments, so initially the market liberalization had only minor effects on competition. The forces that actually led to competition were the Internet and the technological evolution in computer science.

The computational capacity per unit volume of silicon has doubled approximately once every 18 months for the last 30 years (Moore's law). The amount of software available on computers has grown even faster. This allows us to construct more complex applications that are able to perform tasks that were impossible just a few years ago.

As explained in Section 2.3, the Internet has altered the business model of telecommunications in a different and unexpected way. Prior to the Internet, the telecom business was managed by companies consisting of a single vertical structure offering access, transport, services, and even terminal equipment in a single subscription. The Internet has changed this business model entirely by splitting the business into independent parts. Operators called Internet service providers (ISPs) offer access and transport of bits while independent content and application providers (ASPs) offer the services. Of course, one company may be an ISP and an ASP at the same time, but being an ISP or an ASP are nevertheless two different business models.

Voice and video streaming on Internet protocol (IP) removes the last traditional services from the telecom and broadcast providers. The Internet allows anyone owning a computer and a Web camera to produce television programs and distribute them to anyone who cares to view them. A lot of people do this already, and, deeming from the number of hits on various Web sites containing homemade video films, these services are popular.

Despite of this complex evolution, the basic telecommunications technologies such as switching, access, multiplexing, wireless communication, and synchronization are almost unchanged. Moore's law has allowed us to implement old ideas such as code division multiple access (or spread spectrum). When the concept was studied during the late 1940s and early 1950s, devices did not exist on which the idea could be implemented. It took more than 50 years before it was feasible to implement the idea on a large scale. The first real-time simulations of the modulator, the fading radio channel, and the demodulator of global system for mobile communications (GSM) had to be run on a Cray computer. The simulation of a sample lasting 4 seconds took about 1 hour on a workstation. This was in 1986. In 1991—thanks to Moore's law—the modulator and demodulator were contained in small handheld GSM phones.

When I see the specification of an entirely new system concept, I often get the feeling that I have seen it before. The reason, of course, is that the new concept makes use of methods and solutions that, possibly in a slightly different guise, have been used in previous systems. This does not imply that the system does not include entirely new ideas that have never before been exploited. Every idea has a first time. After that the idea may be reapplied, altered, expanded, and so on in order to construct new concepts. Understanding this dynamics of system design is important: if you can reuse something that has been done before, you will save time and sometimes even improve reliability of the new system.

When I took over the course on access and transport networks at the Norwegian University of Science and Technology (NTNU) 5 years ago, I found that the students knew much about how concrete systems were composed and functioned but did not understand how the same technology was reused in different systems. For this reason, I changed the focus of the course away from the description of systems such as asynchronous transfer mode (ATM), GSM, universal mobile telecommunications service (UMTS), and Ethernet to the description of the baseline technologies that were used in the design of these systems. The response from the students—at least the most clever ones—was encouraging. Having participated in standardization and development of systems for almost 40 years, I could also show in a rather convincing way how we had "stolen" ideas from previous designs and put them together in new ways. GSM is almost a compendium in this way of working.

The present text has been developed, in particular, from dialogues with my students. They have offered many suggestions concerning what is important and what can be excluded from the text. Therefore my foremost acknowledgments go all these students.

I am also grateful for many suggestions from colleagues in Telenor and NTNU concerning the contents of this book.

My final acknowledgments go to all the colleagues I have had during more than 30 years of international standardization and system development. For me, this represented a vast arena of knowledge that is the basis for this book!

CHAPTER 1

Introduction

1.1 Evolution of Telecommunications

There are two particular events that have changed telecommunications during the last 25 years. These are the introduction of automatic mobile communications around 1980 and the commercialization of the Internet in the early 1990s. The evolution is illustrated in Figure 1.1.

The Internet gradually replaced the telex service during the 1990s. The telex service offered a method by which text could be transferred between teletypewriters. The system operated at a speed of 50 bits per second (bps) and each symbol consisted of five information bits, one start bit, and one and a half stop bit—or 7.5 bits altogether. The telegraph service using the Morse alphabet lasted until 2000 because the service then was no longer mandatory for ships in international waters by the Safety of Life at Sea (SOLAS) convention of the United Nations. The service is now entirely replaced by the more reliable maritime satellite services, as well as maritime VHF and HF telephony.

The telecommunications operators developed the packet data transmission service called X.25 [named after the International Telecommunication Union (ITU) recommendation where the service is specified]. This service was replaced by the Internet during the 1990s, even though the telecommunications operators had invested large sums in the implementation of the service. The Internet was a cheap alternative to X.25 that moved the control of the service away from the

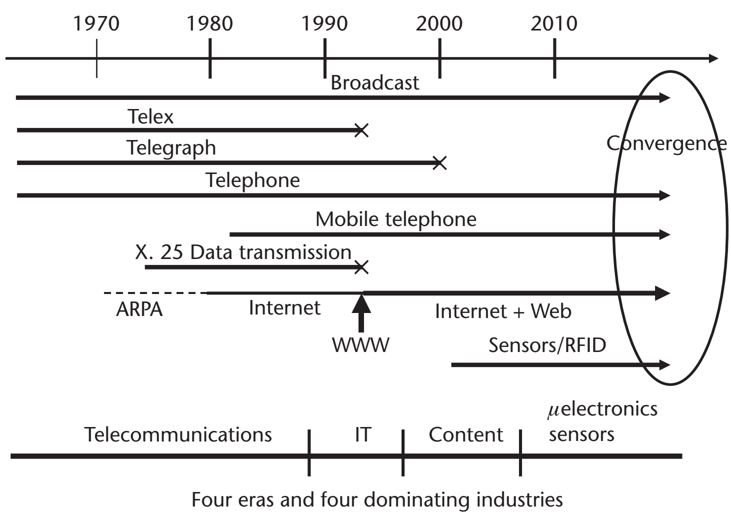

Figure 1.1 Evolution of telecommunications.

bureaucratic telecommunications operators. With the Internet, the users can configure their own services.

One of the most important evolutions taking place at the moment is the expansion of sensor technology, including the radio frequency identification (RFID). It is expected that the microelectronic industry will take the lead in the future evolution of telecommunications. One reason for this is that it is estimated that there are more than 1,000 times as many autonomic devices containing central processing units (CPUs) as there are personal computers, databases, servers, and mainframe computers. The volume of autonomous machine-machine interactions increases rapidly and is expected to constitute a large part of the future telecommunications traffic, both locally and remotely. Some characteristics of the new traffic may include support of frequent and very short transactions, micropayment for use of computation facilities (grids and agent networks), and information security related to anonymity, nonrepudiation, global access control, and nondisclosure of processing algorithms in an open processing environment.

Until about 1985, the telecommunications operators were in charge of the telecommunications business. The business was then mainly concerned with telephony and broadcast. For the next 10 years the information technology industry determined much of what was taking place in telecommunications. During these years, data communication rose and matured. The Web allowed everyone to create and distribute content. This put the content industry in the driver's seat from about 1995. The content industry has changed telecommunications from being a pure telephone system (or person-to-person interaction) to becoming a system that supports all kinds of data communication and, in particular, dissemination of content and information (person-machine interaction). Now the new role of machine-machine interactions enabled by the microelectronic industry may shape the telecommunications industry further.

Still, there are three separate telecommunications networks:

- The telephone network supporting fixed and mobile telephone services;
- Broadcast networks;
- The Internet.

This situation is about to change.

Since the early 1970s, the telecommunications operators have been studying different ways in which all telecommunication services could be supported by a single digital network. The first attempt was the integrated services digital network (ISDN) developed during the late 1970s and the early 1980s. This attempt failed because the ISDN is a circuit-switched network not capable of incorporating packet-switched data communication. However, the ISDN specified how subscribers can be connected to the transport network on a digital access circuit supporting all types of digital services, including Internet access, and allowing several types of terminals to share the same subscriber line.

ATM was developed during the late 1980s and the early 1990s in order to support any mixture of circuit-switched and packet-switched communications in a single system. ATM integrates all services in a single network, but it never became a success because it cannot compete with the Internet in terms of switching costs.

Furthermore, asymmetric digital subscriber line (ADSL) offers sufficient bandwidth on the user access so that the ATM technology also became too expensive on the subscriber line. If large bandwidth is required in the access network, for example, to a local area network, this can simply be supported by an optical link and standard Internet switching both in the LAN and in the network. A separate technology such as ATM is not required for this purpose.

The evolution taking place now is that service integration is finally being achieved by merging all data services, telephone services, and broadcast services in the Internet, as illustrated in Figure 1.2. The connectionless IP network is capable of offering reliable and high-quality, real-time services. This has resulted in voice-over-IP (VoIP) and video-over-IP services. In addition, the 3G mobile network evolves toward an all-IP version, where all information is sent as IP packets on the radio path. This evolution will have deep impact on the telecommunications business as we shall see in Section 2.3.

1.2 What Is Important Knowledge: Generic Technologies or Detailed System Overviews?

One question is as follows: do we need to understand all the basic technologies of telecommunications since the convergence is leading toward a simple network? The answer is affirmative because the technologies are not becoming obsolete even if some of the earlier networks are being replaced by the Internet. This leads directly to the motivation for writing a text that focuses directly on the technologies, thereby shoving the system knowledge to the background.

What then about the details and the functioning of systems in actual use?

A book on access and transport networks can, of course, describe each system in terms of architecture, design, and details concerning protocols and information exchange. In other words, the focus may be on the detailed description of each system: how it is made, how it works, and what it does. Internet, ISDNs, ATM networks, 2G land mobile systems, 3G land mobile systems, and wireless local area network (WLAN) systems may all be described separately. Systems that are still on the drawing board may also be included in order to avoid overlooking a future evolution that may come.

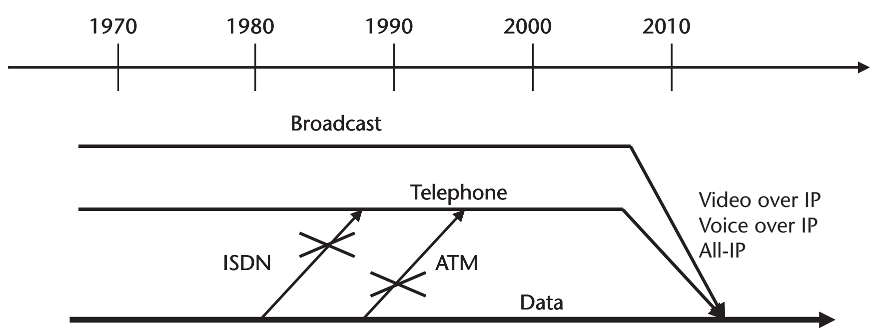

Figure 1.2 Convergence.

One problem with this approach is that systems are replaced by new systems. A few years ago, the ATM technology was a central issue in a course focusing on actual systems. Today, detailed knowledge of ATM must be regarded as rather peripheral, though the technology is still used. GSM is still an important topic because the system is in widespread use all over the world, but in a few years time GSM will become obsolete and, therefore, no longer of general interest.

Having taken part in the early phases of the development of several complex systems such as maritime satellite communications, automatic land mobile systems, and intelligent networks, my experience is that the most important element leading toward successful design is the understanding of the basic technologies that may or may not be useful in the new system. It is not always evident that the same basic technology is often applied in different and unrelated systems.

For the system designer, it is important to understand how a technology can be reused fully or partly in a new design and when an entirely new approach must be found. Therefore, the focus I prefer is to consider the basic technologies rather than the systems in which they are used. The systems, or rather particular features or components of the system, are described in order to illustrate how a particular technology may be exploited in order to achieve a certain result.

A technology that is 30 to 50 years old often appears in current system designs. Code division multiple access (CDMA) is a good example. The detailed mathematical description of direct sequence and fast frequency hopping CDMA was fully developed during the 1950s. At that time, the technology was usually known as spread spectrum multiple access (SSMA). The principles had also been demonstrated experimentally. However, it is just recently that this technology has become mature for large-scale production (3G mobile systems). The reason is the tremendous amount of computation required for detection and synchronization of such signals. This is now possible in even small devices such as mobile phones.

The reason that a particular advanced technology is not applied is often commercial rather than technical: the technology may simply be too expensive. One example is the smart-house technologies that require remote control and sensing of room temperature, which can thereby reduce heating expenses. The technology is simple and has been available for more than 20 years but the equipment at the user site has been to expensive compared to the reduction of the electricity bill.

It took more than 20 years before public key infrastructure (PKI) and electronic signature became commercially feasible. The way in which PKI can be implemented has been known since the early 1980s. The first time I heard about trusted third parties and public key escrows was in the early 1980s. In the early 1990s, one hot item—even resulting in Ph.D.s—was the discussion concerning who had enough trust to own such devices. Still it took 10 years before the first PKI infrastructure was realized. The question of trust is still unresolved.

WiMax offers an alternative implementation of the fixed subscriber line. This is a modern realization of a radio access technology being studied and tested during the 1970s. However, it soon became evident that the technology was too expensive at that time compared to alternative subscriber lines. In many implementations, the WiMax technology is still too expensive, even for a new operator that wishes to establish an independent access network. The alternative of leasing access from an incumbent operator may be cheaper because the telecommunications regulatory authorities are fixing maximum prices for so called local loop unbundling (LLUB) that quite often are cheaper than building a new access.

Random access, including the control procedures applied in WLAN and Ethernet, was analyzed by Kleinrock and others some 30 years ago. All the "modern" switching and multiplexing methods of digital signals were developed almost 40 years ago, and some of these technologies are recently being extended to optical switching and multiplexing. The IP technology has been with us for the last 35 years. Automatic (or cellular) mobile communication with full roaming capabilities and handover was put into operation in 1981.

What is really new is that the perpetual evolution in microminiaturization and computing enables us to implement more and more complex systems.

Furthermore, particular systems such as GSM and telephone switching are at one stage becoming obsolete and replaced by new (but not necessarily better) systems. The detailed knowledge of these systems is thus only of limited value. However, the technology on which they are based does not become obsolete but may be reused in entirely different systems. Frequency division multiplexing (FDM) has been regarded as an obsolete technology in the telecommunications network for a long time. The technology is now reappearing in a slightly new guise in optical networks, where it is called wavelength division multiplexing (WDM), and in broadband access networks, where FDM is used to increase the effective bandwidth that can be supported by the twisted pair (direct multitone ADSL).

ATM is a technology that is disappearing from the network and as such should not be of interest in a course like this. However, ATM contains some features that may be reused in new designs. Two such features are the use of error detection to synchronize the data stream and the use of length indicators to multiplex different information streams into a common cell structure. The latter method is used in several systems, but the way it is done in ATM is easier to describe and simpler to understand.

For these reasons I have focused on the basic technologies applied in access and transport networks rather than the actual systems. Actual systems are used as examples in order to show how the technology is used in particular circumstances.

The description of particular systems is contained in numerous textbooks and standards documents, and the reader is referred to such literature in order to study the details of these systems.

1.3 Composition of the Text

The book consists of 10 chapters as follows.

Chapter 2 contains general definitions and explains some particular features of telecommunications systems, such as the distinction between intelligent networks (ISDNs) and stupid networks (Internets), domain structures, and overlay access and virtual networks. The chapter is also concerned with problems such as real-time operation, heterogeneity, backward compatibility, and standardization.

Chapter 3 is about synchronization. One important item is the description of the phase-locked loop (PLL). The PLL is one of the most important components in digital networks and is used for bit timing acquisition, carrier acquisition and coherent demodulation, frequency synthesis, and many other applications. PLLs are included in multiplexing equipment, signal regenerators, satellite systems, radio relays, land mobile terminals, and so on. A general description of the loop is contained in the main text. The loop mathematics and construction details of analog loops are contained in the Appendix.

A large number of applications of synchronization are described. These include the interconnection of synchronous and plesiochronous networks, synchronization in ATM where the error correction mechanism is used for maintaining cell synchronism, synchronization of TDMA satellite systems, timing advance in GSM, and signal detection in WLANs.

Chapter 4 describes several multiplexing methods used for static and statistical multiplexing. Static multiplexing includes frequency division multiplexing, time division multiplexing, and the synchronous digital hierarchy (SDH). Statistical multiplexing methods include systems with constant length envelope (ATM), use of length indicators (also used in ATM), and variable length envelopes using flag delimitation and transparency mechanisms.

Chapter 5 is concerned with multiple access; that is, techniques that allow several sources to share a common medium. The basic methods of frequency division, time division, and code division multiple access are explained in detail. The chapter also contains an introduction to random access explaining how the method is applied in satellite systems, GSM, Ethernet, and WLAN. One important part is concerned with the stability of random access channels and the methods that can be applied to avoid channel saturation. The particular methods used in WLAN, Ethernet, the Internet, and other systems are explained.

Chapter 6 is concerned with switching systems. The chapter consists of two parts. The first part is concerned with network aspects of switched networks explaining how routing and switching takes place in circuit switched networks (ISDN), connection-oriented packet switched networks (ATM), and connectionless packet switched networks (Internet). Features such as number analysis are also explained in relation to the different technologies.

The second part describes in general terms how space and time division switches function. Then a more detailed description of particular switching networks is provided, including the general Clos-Benes network and the application of binary switching matrices in fast switches for ATM and optical networks.

Chapter 7 contains the basic elements of protocol theory. Protocol theory is basic knowledge required for understanding signaling systems and data transfer protocols. The chapter contains three examples:

- Embedding and tunneling in the Internet in order to support mobile IP and particular network related protocols.
- The structure of Signaling System No. 7 (SS7). This signaling system is used in the ISDN/telephone network and in mobile networks in order to support the interaction between the different entities making up the mobile network.
- The protocol structure of GSM shows how complex protocol stacks may be required in order to support a large number of functions in the different environments in which the system is embedded.

Chapter 8 describes public land mobile systems. The chapter contains both general information such as radio propagation phenomena and generic network architecture, as well as details concerning GSM (2G), general packet radio service (GPRS) (2.5G), and UMTS (3G). One particularly important goal is to show similarities and differences between these systems. The new evolution toward software radio, led by the terminal manufacturers, is also explained.

Chapter 9 is concerned with line-of-sight radio communication systems. The chapter contains two brief sections on WiMax and radio relays, respectively. WiMax is a technology based on the Institution of Electrical and Electronics Engineers (IEEE) WLAN standards that is about to be implemented in the access network. WiMax may change the telecommunications market entirely. The major part of the chapter is concerned with fixed and mobile satellite communication in the transport network and the access network. Though optical fibers have replaced several intercontinental satellite systems, there are areas where satellite communication cannot be replaced easily by other technologies.

Chapter 10 describes briefly the various components of optical communication systems. Several of these components and systems are still not commercially available because of cost, size, and manufacturing complexity.

CHAPTER 2

Networks and Services

2.1 Access, Transport, and Platform

The basic composition of telecommunications networks is shown in Figure 2.1. This configuration applies to all types of networks: the telephone network, the ISDN, data networks, broadband networks, the Internet, and so on. There is no structural difference between networks for different purposes at this level of abstraction. The difference is apparent when we consider construction details. The simple subdivision consists of three elements:

- The *access network* connects the terminals or users to the transport network and transports bits across the user-network interface. The access may be complex and contain much functionality, such as in mobile systems, or it may be simple, such as in the fixed network. The access network may consist of several technologies in tandem (e.g., Ethernet, WiMax, and local optical fiber).
- The *transport network* connects one access to another access via switching devices and other machines in the platform. The main purpose of the transport network is to transfer bits between two access networks.
- The *platform* routes the call from the origin to the destination. The software and hardware in the network and in the terminals cooperate in performing the processing of services and applications. The need for processing in the network depends on whether the network is stupid or intelligent (we'll discuss these terms later) and the types of services offered.

The network consists of routing devices (e.g., ISDN exchanges in the telephone network or routers in the Internet) that offer three types of services: transport of bits, routing, and mobility.

2.1.1 Transport of Bits

This is the main function of the access network and the transport network. The technologies used for transport of bits are usually different in the access network and in the transport network. The most common technology used in the transport network is optical fibers with bandwidths of several gigabits per second (Gbps). But other technologies offering less bandwidth— coaxial cables, radio relays, and satellites—are also used extensively, though some of them are gradually being taken out of use and replaced by optical fibers where possible.

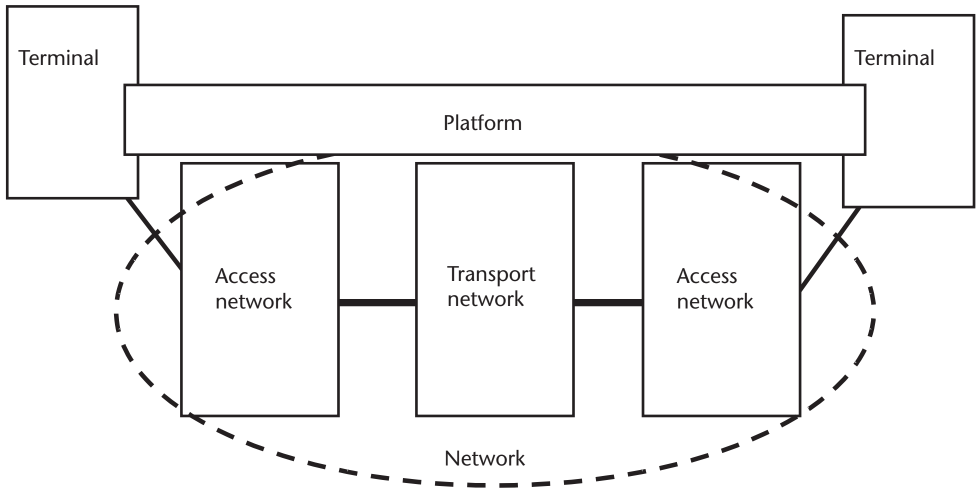

Figure 2.1 Elements of the telecommunications networks.

The technologies used in the access network are much more diverse: twisted pairs, coaxial cables, optical fibers, WiMax, GSM, UMTS, WLAN, Bluetooth, satellite systems, and several other technologies. Some of these technologies are considered in more detail in other chapters.

2.1.2 Routing

Routing is the core function of the switching process. The routing function selects the output of the switch on which a datagram or a telephone call is to be forwarded in order to reach the destination. How this is done is explained in Chapter 6.

With the Internet, routing is the functionality of IP and is a rather simple process requiring simple switches. With ISDN, the routing requires complex switching functions. One reason that the Internet and ISDN are different is that IP is a connectionless protocol that applies statistical switching, while ISDN is a connection-oriented protocol requiring separate signaling functions for establishment, management, and release of connections.

Switching also includes several processes in addition to routing that can be applied to the call: access control such as barring access to or from certain users, nondisclosure of source addresses, and measuring call data for determining charges and for building traffic statistics. These processes are generally more complex with ISDN than the Internet. This is another reason why ISDN exchanges are so complex.

2.1.3 Mobility

Mobility allows terminals or users to roam between different access points. Mobility is offered by GSM, UMTS, WLAN, and mobile IP. GSM and UMTS offer continuous or nondisruptive mobility, while mobile IP offers discrete mobility. WLAN offers something in between: continuous mobility within a WLAN zone and discrete mobility between zones. Personal computers, personal digital assistants (PDAs), smart phones, and mobile phones are equipped for several access technologies.

Personal computers support any selection of access interfaces: cable connection, GSM/GPRS/UMTS, WLAN, Bluetooth, and WiMax.

2.2 Types of Networks

2.2.1 Transport (or Backbone) Network

The most important technologies applied in the transport network are as follows:

- *Optical fibers*. These systems have replaced most other long distance systems during the last few years.
- *Geostationary satellites*. Some of the largest systems have been (or are about to be) replaced by optical systems because the latter are cheaper, offer larger bandwidth, and are more reliable than satellite systems. However, the satellite systems are still used on several intercontinental routes where the traffic is low or where it is too expensive to install optical fiber. In addition, there are several domestic systems covering areas that are otherwise impossible to reach or where alternative systems are too expensive (such as Australia, Canada, Indonesia, and Brazil).
- *Radio relays*. The radio relay systems are in general cheap, reliable, and easy to establish or rearrange. The bandwidth can be any multiplex rate between 2 Mbps and 640 Mbps. These systems are used in applications such as feeder links in the local part of the network, in mountainous areas where it is too expensive to provide optical fiber systems, and in earthquake areas since it is easy to reestablish the link after an earthquake.

These systems are described in later chapters.

2.2.2 Access Networks

The access networks can be divided into three classes: public fixed access systems, public mobile access systems, and local area access systems.

Public fixed access systems include the following:

- Twisted pairs (or copper lines) are by far the most common access system in the fixed network. The twisted pair has a very long technical lifetime (in excess of 50 years). The twisted pair supports broadband up to 10 Mbps in the form of ADSL and other digital subscriber line of type x (xDSL)[1] technologies.
- Optical fibers are installed in areas where twisted pairs do not exist and are also used to replace twisted pairs elsewhere. These systems are still rather expensive, which is one reason for the survival of the "obsolete" twisted pair technology.

1. DSL stands for digital subscriber line. The A in ADSL stands for asymmetric because different bandwidth is allocated to the uplink (from the user) and the downlink (toward the user).

- Coaxial cables are still abundant in cable television systems, though more and more systems are being replaced by optical fiber. However, coaxial cables have a very long technical lifetime (50 years or more), so there must be strong commercial arguments for replacing them by fibers—wider bandwidth and support of duplex services (telephony and the Internet) are such arguments. Coaxial cable can support a bandwidth in excess of 200 Mbps.
- Broadcast satellite systems can offer broadband Internet services on the downlink to the user. Narrowband Internet may be used in the opposite direction for providing full duplex services.
- WiMax is a fixed point-to-point radio system (also capable of mobile communication) doing much the same job as twisted pairs but with a higher bandwidth (100 Mbps). WiMax is still an expensive technology, though the equipment cost is dropping. WiMax has become competitive in several applications, such as connecting WLAN hotspots to the transport network and as access network in new housing regions.
- Electricity modems are used to provide telecommunications services over the local electrical grid. Bandwidth of a few megabits per second is possible. However, the technology is still expensive and is used only rarely.

Public mobile access systems comprise the following:

- GSM, GPRS, and UMTS (3G) land mobile communications. Competing technologies based to the UMTS system exists in the United States and Japan. These systems are also introduced as competitors to 3G in Europe—for example, in the 450-MHz band previously used for the Nordic Mobile Telephone System (NMT) and other early mobile systems.
- Maritime and aeronautical satellite systems employing four geostationary satellites to cover all ocean areas and flight routes except the Polar Regions (i.e., coverage between latitudes approximately +70 degrees and –70 degrees). The same satellites are also used for land mobile communication to remote areas, for relief and rescue operations, and for expeditious establishment of broadband access systems (e.g., on-the-spot television reporting).
- Low orbit satellite (LEO) systems have been tried (Iridium, Globalstar, and other systems) but were not competitive with GSM or other public land mobile systems for general mobile communication. The Iridium consortium and Globalstar went bankrupt after having commenced full service in 1999 and 2003, respectively. However, the satellites were later sold to other companies reestablishing the services. The systems offer telecommunications to governments, the oil industry, scientific explorations, relief operations, and travelers. By the end of 2005, Iridium had 142,000 subscribers. The Teledesic, originally planned with 824 LEO satellites, later downscaled to about 300 satellites, offering “IP in the sky,” was never realized because the company was scared off by the bankruptcies of the other companies. However, Teledesic is still regarded as an alternative to the terrestrial Internet in the future.

Local area access networks comprise the following:

- Ethernet is a fixed local area access network using twisted pair, coaxial cable, or optical fiber as transmission medium.
- Wireless LANs in one shape or another (e.g., IEEE 802.11) offering short distance communication are emerging rapidly in the unlicensed frequency band around 2.4 GHz. Several related technologies providing larger bandwidth are under development.
- VSAT systems are local area networks interconnected by geostationary satellites.
- Bluetooth interconnects devices locally in addition to offering a communication port to external networks (e.g., a WLAN or an Ethernet).

2.3 Stupid and Intelligent Networks[2]

2.3.1 Concept

The Internet is a stupid network. By this we mean that the Internet is offering very few services to the users. The Internet offers primarily routing and transfer of datagrams. In addition, the Internet manages multicast addressing and routing, allocates bandwidth, and supports mobile IP (tunneling). The routing and delivery of datagrams is based on "best effort"; that is, there is no guarantee that a datagram ever reaches the destination or is lost because of congestion in the network.

With the Internet, all intelligence in terms of myriads of applications is in the hosts (PCs, databases, servers, sensors, or other computing devices). The network is stupid but the terminal is intelligent. In other words, the Internet contains a stupid core but an intelligent periphery.

On the contrary, the ISDN/telephone network is intelligent, while most of the terminals connected to it are stupid, supporting very few applications. The network is offering several intelligent services, such as barring of incoming and outgoing calls based on time, cost, origin or destination, call redirection, call waiting, recall services, premium rate charging, toll-free services, alternate charged party, charge sharing, nongeographic routing, centralized queuing, distributed call desk, auto-answering services, voice mail, conditional routing, and many more. Some of these services are performed in separate devices called intelligent nodes. The core of the ISDN/telephone network is intelligent, while the periphery is stupid.

GSM, GPRS, and UMTS offer services to both stupid and intelligent terminals and represent a transition between intelligent and stupid networks. The mobile terminal contains one or more CPUs, and therefore the mobile phones belong to the category of intelligent terminals since several services may be designed in the terminal rather than in the network. As long as these networks offer telephone services in

2. These terms were introduced in the Telecommunications Information Networking Architecture (TINA) project in order to distinguish between two fundamentally different types of networks. The term stupid network was introduced at about the same time by David Isenberg, "The Rise of the Stupid Network," *Computer Telephony,* August 1997.

its core, the mobile network contains both an intelligent core supporting supplementary services for telephony and an intelligent periphery supporting data transmission and additional capabilities. The all-IP 3G network (delivery 5 of the UMTS specification) offers only IP-based communication (including IP telephony) over the access so that the core of this network no longer will contain intelligent functions.

VoIP or IP telephony in the Internet is a telephone service where the intelligent services of the network can no longer be supported unless additional functionality is added to the Internet and the Internet protocols. The owner of the IP network (or ISP) may offer particular handling of IP telephone calls similar to that of intelligent network nodes in separate servers owned by the ISP. However, these intelligent services are services above the demarcation line described next and thus belong to the periphery of the network. The intelligent services are implemented in servers and databases.

In summary, a stupid network has a stupid core supporting an intelligent periphery (the Internet); an intelligent network has an intelligent core supporting a stupid periphery (the telephone network). The trend is that all telecommunications networks are developing in the direction of stupid networks. The era of intelligent networks may soon be over.

2.3.2 A Note on the Protocol Structure of the Internet

The protocol structure of the Internet is explained in this section so that the main part of the text can be understood without referring to other literature where the Internet protocols are described (see also Chapter 7 for an introduction to protocol theory). This description contains only the elements and details of the protocol that are required to understand the main arguments presented in the following discussion.

The protocol structure of the Internet is shown in Figure 2.2. The figure shows two computers containing application software communicating over a network consisting of two routers (R) in order to perform a common task (for example, a Web search where one computer contains the browser and the other computer contains the search engine). The protocol only ensures that the transaction between the two computers can take place.

The computers are interconnected by a protocol consisting of several layers, where each layer is a protocol in its own right. The layers are as follows.

The *lower layers* (link in the figure) support reliable transfer of bits between one node and the next. (The link protocol may consist of several layers. This point is not important here.) The link protocol may be different on different links depending on the characteristics of the transmission medium. The link protocols are different in, for example, optical core networks, GSM, UMTS, WLAN, and fixed broadband accesses. The link only analyzes the link header(s) and not the information contained in the information field. The information field of the link layer contains the IP datagram.

IP is a *network protocol* whose primary task is to route bits from one terminal to another. The IP protocol may be identical across the whole connection (e.g., only IP version 4) or consist of sections with IP versions 4 and 6 in tandem. This can be done since the IP header is analyzed at each router so that the router can determine which actions should be taken, including protocol conversion and tunneling (see Section

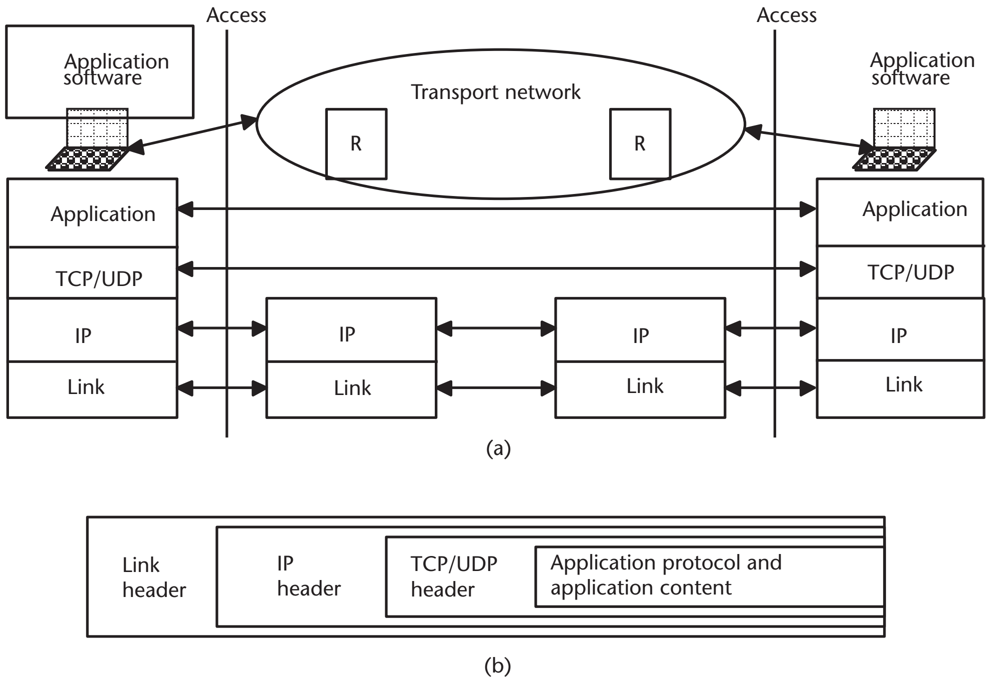

Figure 2.2 (a, b) Protocol hierarchy in the Internet.

7.7.2 for tunneling of IPv6 across IPv4 networks), when the IP packet is forwarded on the next link.

When forwarding the packet, the router then creates a new IP header containing information inserted by the router (e.g., a new value of the time-to-live parameter) and header parameters copied from the received packet (e.g., IP addresses). The router does not analyze the content of the information field of the IP packet.[3] However, the IP packet contains a parameter identifying which protocol the information field contains—another IP,[4] transmission control protocol (TCP), or user datagram protocol (UDP) (in IP version 4 this is the Protocol field; in IP version 6 this information is contained in the Next Header field—see the specialized IP literature for further details). This information is required by the receiving terminal so that it can identify which software must be activated in order to interpret the content in the information field (e.g., the software required for handling TCP or the simpler software for handling UDP). The router may use the next header information so as to handle datagrams containing TCP or UDP differently (routing selection and buffer

3. Except the port number contained in the TCP/UDP header (shared address). However, this is not important for the discussion that follows, though it allows the ISP to have some knowledge of what the protocol contains. The shared address field allows address extension of the IP number by using one of the port addresses for the extended IP number (see the specialized IP literature for how this is done).
4. In order to support mobile IP or security (IPsec) the information field contains another IP protocol (see Section 7.7.2). The embedded IP protocol may then contain a Protocol/Next Header field indicating that the information field contains a TCP or UDP header or even another IP header (e.g., for embedding an encrypted datagram).

priority), since UDP is used for real-time services while TCP is used for data transmission where there is no timing constraint. The IP packet is embedded in the information field of the link protocol, as shown in Figure 2.2(b).

The layer above the network protocol (IP) is called the *transport layer*. The header of the transport layer protocol contains information (with the possible exception of the port address—see the footnote) that is only read by the terminals and ignored by the routers. The most common protocols are TCP supporting a connection-oriented data transmission service between the two terminals and UDP supporting connectionless transfer of real-time information (voice and video). TCP and UDP contain a parameter called *port number*. The port number identifies the type of information (application protocol) included in the information field (e.g., port 80 for http and port 23 for telnet). Even if we know the port number, we may still not know the actual applications the port supports. Port 80 is used for diverse services such as Web search, Web management, XML Web services, and even some types of Internet telephony. Therefore, the actual service supported cannot be identified from information contained in the TCP/UDP header alone.

The *application protocol* is designed for managing a particular application, such as Web search. In this particular case, http is used as application protocol. Application protocols exist for all types of services and applications offered on the Internet (e-mail, file transfer, remote procedure call, and so on). Note that the application protocol is not part of the application but assists the application in transferring the application content (information or commands) across the network.

The application software (or the middleware if it exists) contains instructions that request the terminal to initiate the protocol stack whenever a remote transfer is required.

In this context, a terminal can be any type of equipment containing a CPU: PC, mainframe computer, server, database, mobile phone, sensor, actuator, printer, smart card, and so on. While there are almost 2 billion (2×10^9) PCs in the world today, it is estimated that there are more than 1 trillion (10^{12}) devices (mainly sensors and actuators) satisfying the stated definition of a terminal. Most of these devices are either directly or indirectly connected to the Internet.

2.3.3 The Line of Demarcation Between Network and Application in the Internet

The Internet can be divided into two parts by a demarcation line, as shown in Figure 2.3. Below the line we have the network consisting of routers, the access, and the IP card in the terminal. Above the line we have the applications or the software (including the application protocol) running in the terminals. The demarcation line is in fact the transport protocol (TCP or UDP).

On the two sides of the demarcation line, the business and the user charging models are completely different. The telecom operators and ISP reside below the line. They own routers, cables, and support systems for running the network. This is the traditional telecommunications business. In principle, the operator/ISP may use traditional charging, such as charging the customer for being connected (access charge); basing the charge on volume indicators such as the number of bits or IP packets sent or received; basing the charge on the actual bandwidth used; charging the customer for the duration of the connection; charging various content (voice,

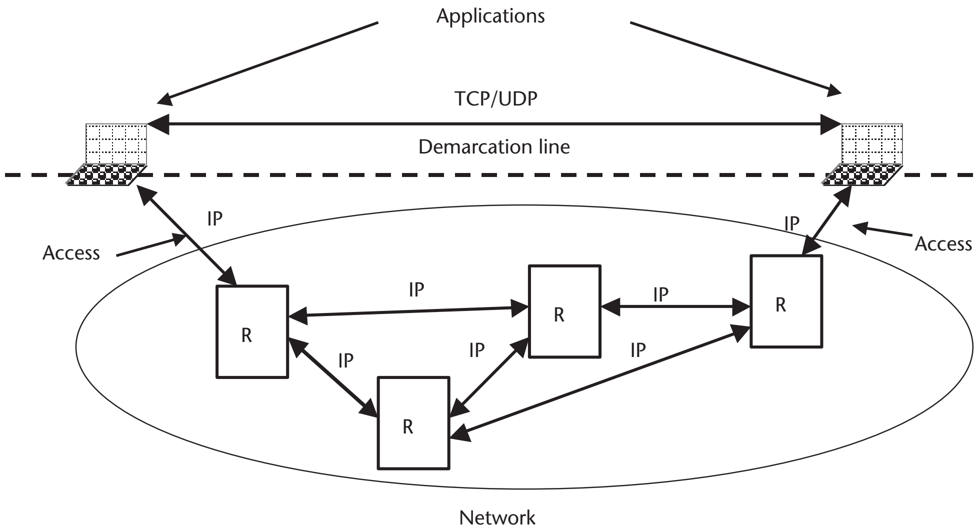

Figure 2.3 Separation of Internet functionality.

picture, video, data files) differently; or letting the usage of the network be free because the operator/ISP earns money from other sources such as advertising. However, as we shall see, the demarcation line makes all this (except levying access charges) difficult or perhaps impossible.

Above the line, we have providers of all types of services and information that only need a medium offering sufficient bandwidth over which information can be transferred: films, music, Internet telephony, Web search, Web conferences, electronic newspapers, e-banking, e-mail, e-commerce, remote sensing and control, and so on. The transfer medium just happens to be the IP network. Any network that offers bandwidth adaptable to the data rate of the source could have done the job—this was just why ATM was developed. However, IP did it cheaper, so therefore the effort to replace the transport network with ATM failed.

The problem for the service and information provider is getting paid for the service, the application, or the information. This problem has turned out to be rather difficult, since the range of services and applications is large and the user's willingness to pay for an application is rather unpredictable.

For some of these services, charges may be hidden in other fees (e-banking fees, credit card fees) or included in the price of the goods (e-commerce); other services are financed by advertisements: selling customer databases and information on user behavior (Skype); in still other cases, the service is free of charge because it supports a complementary service on which the provider earns money (Google); finally, the customer is not willing to pay for some services (electronic newspaper).

The model in Figure 2.3 can also be drawn as shown in Figure 2.4, where the Internet is split up into three independent networks: the IP network containing routers (R) and the IP interface in the terminal (circles); a dynamic network consisting of terminals or other devices—servers (S), personal computers (PC), databases (D), or any other computing device—that communicate over TCP/UDP; and an even more dynamic and complex network of interacting application software (A).

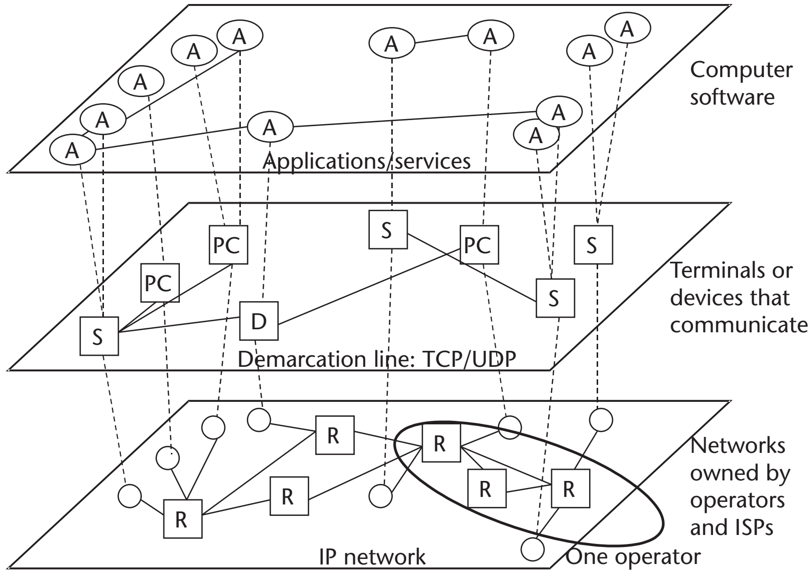

Figure 2.4 The Internet is three networks and not just one.

Network operators and ISPs reside in the lower plane. The actors in the plane in the middle are everyone (persons or firms or organizations) owning a terminal. In the upper plane, we find all kinds of people, firms, and organizations selling or giving away any type of information, content, service, software, transaction, support, or anything else that can be coded as strings of zeroes and ones.

2.3.4 Network Neutrality

The demarcation line is the basis for the notion called *network neutrality*—or as cartoonist Peter Steiner has put it: "on the Internet, nobody knows you're a dog." The applications residing above the demarcation line are egalitarian in that everybody's information packets are treated in the same way by the IP network. This allows business models where everyone may create content and distribute it without being treated differently depending upon the type of content being distributed and what the provider is—be it a broadcast company or an entrepreneur working out of the garage. However, this causes problems for the network providers because old business model based on volume and time charging may no longer be feasible.

Network neutrality encompasses the following four freedoms (of course, subject to legal restrictions) for the users of the Internet:

- Freedom to access content on the network (i.e., access to the information can only be regulated by the owner of the information and not by an ISP or another third party not operating on behalf of the owner of the information);
- Freedom to run applications of any kind alone or together with other users;
- Freedom to attach any hardware to the network (e.g., routers, servers, PCs) that satisfies the Internet specifications;

- Freedom to obtain information about all services and electronic goods available on the network.

Network neutrality is, of course, subject to political debate. The supporters of network neutrality claim that the principle is in favor of competitive market evolution, since many applications and content provides can operate on the same arena and thus increase the total national revenues generated by the network. Network neutrality also favors innovation, experimentation, and provision of services that are too small and too specialized to be considered seriously by the large ASPs.

The opponents claim that network neutrality is bad for the network and the national economy, since the revenues from network operation will become too small to support the future evolution of the network. Therefore, it is claimed that the price of the access should depend on the quality of service (QoS) offered by the ISP. Such QoS parameters may include bandwidth, secure delivery of data, real-time operation, privacy and integrity of data, and priority. This includes both fixed access charges and variable charges depending on volume.

The supporters of network neutrality claim that this use of QoS will make the network no longer neutral but favor those who will pay more for the access and thus introduce an unfair competition arena. The opponents claim that the ISP should be entitled to recover their investments by charging for the actual use of network resources.

It is likely that this debate will continue for a long time. There are strong commercial interests among both the supporters and the opponents of network neutrality.

2.3.5 The Commercial Life Below the Demarcation Line

The telephone network is connection oriented. This means that the connection between the communicating parties is set up at the beginning of the call and released at the end of the call. This allows the network operator to count the number of calls made by the user and the duration of each call, and base the charging on these measurements. If different bandwidth and processing (e.g., premium rate charging, toll-free charging, and shared charges) are associated with the call, this may also be taken into account when computing the charge. In this charging model, it does not matter which of the parties are sending the largest or smallest amount of information (or whether or not they are exchanging any information at all): the charges are usually levied against the user initiating the call (where toll-free and shared charges are two exceptions).

On the contrary, this charging principle is not possible with the Internet because the IP network is connectionless, so that all packets must be treated as singular events. There is no way in which the network operator (or ISP) can correlate IP messages in order to determine the start and the end of a transaction, identify which party initiated the transaction, or measure the amount of information exchanged between the communicating parties during the transaction. The only simple charging method is based on subscription. The number of subscriptions reaches maturity when everybody has one or a few subscriptions, each satisfying their overall need for telecommunications. The revenue of the network operator is then independent