Contents

Preface

Our ability to apply space technology to communication needs on earth is necessarily dependent on the ground segment, and the earth stations that comprise it. Little has been documented on the methodologies and practices that ground segment developers employ to design and implement these facilities. Another aspect is the powerful trend toward consumer-style user terminals that bring satellite communications down to an affordable and practical level. This situation was the primary motivation for this handbook, and we were able to draw on over 30 years of development in this field when putting it together.

The book is organized into 11 chapters that discuss the history, requirements, systems engineering, hardware design, and operations and maintenance of earth stations and user terminals. Our approach is to provide a thorough understanding of the technology and implementation issues for the ground segment, rather than delving into copious detail on each element. This is how we could keep the book to a tractable form and allow the reader to learn how to develop specific approaches for fixed, broadcasting, and mobile satellite applications. A cover-to-cover reading will provide a comprehensive review of almost every aspect of the system, considering real-world aspects of earth station design and engineering. From there, the book provides a reference for subsequent analysis and selection of the appropriate design approach. References containing detailed information on specific engineering issues and solutions are included with each chapter.

We begin in Chapter 1 with a brief history of the satellite communication earth station, which is fundamentally a radio station operating at microwave frequencies. We draw heavily from the fields of radio-astronomy (the first dish antennas were actually radio-telescopes), radar, and terrestrial microwave. Basic engineering principles common to all earth stations are contained in Chapters 2 and 3, the latter delving in particular into the RF link budgeting process for geostationary earth orbit (GEO) and non-GEO satellite transmissions. Chapter 4 addresses ground segment requirements for two-way communications services for telephony and VSAT data applications. One-way (broadcast) service requirements are covered in detail in Chapter 5, is by addressing digital video as well as audio and data broadcasting. The data aspects are of particular interest in the context of applying the Internet Protocol over satellite links.

Chapter 6 addresses what has become a vital concern in designing integrated ground segments for digital services—that of baseband architecture. The topic, which defied proper characterization in the past, involves complex data processing, information transfer over a network, multiple access tradeoffs, modulation and coding, and a host of other aspects of creating an automated network environment that satisfies subscribers and other users. We address an overriding concern about implementing competitive architecture that can make money in commercial services. The more traditional topic of earth station RF and equipment design is addressed in detail in Chapter 7, extending from the general subject of gain budgets (EIRP and G/T) to the specifics of the antennas, high power and low noise amplifiers, and up- and down-converters. Also considered in the chapter are important design and performance issues such as group delay, local oscillator stability, and phase noise. Interaction of all of the RF and baseband elements is covered in Chapter 8, which deals with signal impairments and PC software tools for link analysis and end-to-end simulation.

Our next focus is on fixed and mobile user terminals, covered in Chapter 9 from a design and performance standpoint. Also addressed is the development of baseband and modem functions through the principles of the software-designed radio, as well as specific considerations for handheld satellite phones. Design requirements and implementation of major facilities for large earth stations are covered in detail in Chapter 10. We round out the handbook in Chapter 11 with a comprehensive discussion of operations and maintenance principles based on practical experience with many successful projects over the years. This considers both the operational needs of the earth station and network, as well as lessons learned in managing the human side of the equation.

The handbook was designed for working satellite communications engineers and telecommunications specialists who need to consider how to design, implement, and manage the ground segment. It is also appropriate for specialists in RF, baseband, and digital network technology who wish to gain a better understanding of the total system and the interaction of its key elements. Operations and maintenance personnel should find the book useful as an introduction to ground segments and earth stations, and as a handy reference. From an education standpoint, much of the material was used for an engineering short course at UCLA Extension, where it will become the course text.

There is a lot that goes into a book like this, which captures much of a lifetime's experience with the practical side of applying technology to a real need. I would like to express my appreciation to my technical reviewer, Ray Sperber, for his outstanding guidance and input during development of the manuscript. Also, I wish to thank Bill Bazzy, Mark Walsh, and Barbara Lovenvirth of Artech House for their support from start to finish. I am very appreciative of my wife, Cathy, for helping get the project together and assisting me along way—wherever and whenever she could. Without her, there would be no *Ground Segment and Earth Station Handbook*.

Bruce Elbert
Application Strategy Consulting
bruce@applicationstrategy.com

4

Two-Way Communications Service Requirements

The foundation of most telecommunication services is the two-way interactive mode, typified by a telephone conversation between two people. This is, after all, something that we learn from birth, and provides the principle means of exchanging information, views, ideas, and instructions during the normal course of our lives. This section addresses four implementations of satellite communication services: voice networks between fixed points on the ground, narrowband data communications using very small aperture terminals (VSATs), and mobile satellite communications to accomplish in remote regions what cellular and personal communications services/personal communications networks (PCS/PCN) can in developed areas, and broadband services to support or supplement the Internet. Table 4.1 provides a summary of the capabilities of these services as they exist at the time of this writing, while the discussion that follows reviews their characteristics as they relate to the ground segment.

The three types of multiple access (MA) that can be applied to these networks include frequency division multiple access (FDMA), time division multiple access (TDMA), and code division multiple access (CDMA), reviewed in Chapter 2. As reviewed in Table 4.2, each has its merits and issues as far as application in satellite networks. FDMA is appropriate for sharing of spectrum and power within a transponder; it may be used in

Table 4.1

Characteristics of Two-Way Interactive Services Provided by the Ground Segment

Characteristic	Orbit Configuration	Bandwidth per Channel	Multiple Access	Application	Ground Segment Implications
Fixed voice networks	GEO	8 to 64 Kbps	FDMA and TDMA	International telephone trunks to extend the global network, and thin route and rural services in developing regions.	May employ conventional bent-pipe satellites as well as some processing satellites that can reduce earth station size and cost.
VSAT data networks	GEO	64 to 512 Kbps	TDMA and CDMA	Private data networks within countries and regional networks to promote businesses that require greater extent.	Employs partial transponder capacity. Users share resources of the hub and network control.
Mobile satellite communication	GEO, MEO, and LEO	4 to 20 Kbps	FDMA, TDMA, and CDMA	National, regional, and international roaming for mobile telephone and low-speed data, augmenting the global fixed and mobile networks.	Requires a total integrated system with a common air interface. Major operators develop, manufacture, and distribute user terminals.
Interactive broadband	GEO, MEO, and LEO	64 Kbps to 155 Mbps	FDMA and TDMA	Public Internet and multimedia applications, local and backbone.	Requires powerful uplink or high-gain satellite antenna; service to fixed terminals, hubs, and mobile terminals on vehicles, aircraft, and ships.

Table 4.2
Characteristics and Application of Multiple Access Methods in Satellite Communications Ground Segments

Multiple Access Method	Principle Application	Benefits	Issues
FDMA	Sharing of transponder bandwidth and power	Minimum coordination among users	Power sharing, adjacent channel interference, and intermodulation
TDMA	Digital multiple access networks with consistent timing and coordination	Efficient in terms of bandwidth and power usage; high degree of integration with networks	Limited ability to control interference using separate timing or frequencies; requires timing reference
CDMA	High RF interference environments; variable bandwidth and power availability	Can tolerate self and external interference; little coordination among users; low potential to cause interference	Power control; adequate bandwidth for spreading

SCPC DAMA networks and for operating different applications within the same bandwidth. TDMA, on the other hand, requires tight coordination of timing and bandwidth usage, and may be susceptible to interference (as is FDMA). It is beneficial for digital networks because it allows multiple logical channels within the time frame and integrates well with control channels and back-end systems. The third method, CDMA, has been touted for its efficiency in environments characterized by heavy interference from internal and external sources. One benefit is that the transmissions can be adapted to the instantaneous environment, which is often required in a real-world network.

Because of the highly variable nature of the different applications and environments, these methods represent the principle points of departure for the design of appropriate network architectures and rely upon the design and manufacturing capability of industry specialists such as Motorola, Hughes Network Systems, Gilat, Nokia, Ericsson, Alcatel, and NEC. In our experience, the selection of the optimum MA technique relies more on the experience base and economics of the particular supplier and operator as opposed to the ideal technical features that one might analyze on a sheet of paper.

Any architecture that uses a start topology (or those meshes that require connections back to the PSTN or the Internet) will require one or more hub earth stations. This is illustrated by a typical GEO MSS system, shown in Figure 4.1, that has three types of earth stations: the L-band user terminal, the C-band gateway earth station (facilities in three different countries or regions are indicated), and the shared Network Management Center and Subscriber Management System. This particular network does not allow direct UT-to-UT connections as it only offers services between users and the PSTN via the gateways. We see that this architecture offers tight control (vital functions, not to be underestimated) and management of the network, including measurement of usage and billing.

The following sections review some of the detailed requirements for these types of networks that can be served through a satellite communication ground segment. This is not meant as a comprehensive treatment but rather an introduction to this type of investigation.

4.1 Fixed Telephony Voice Networks

Voice services are inherently interactive in nature and so the service must adhere to user expectations and quality objectives. We have as a basis the work done over the past decades at major institutions like Bell Laboratories in the United States and European industrial and intergovernmental

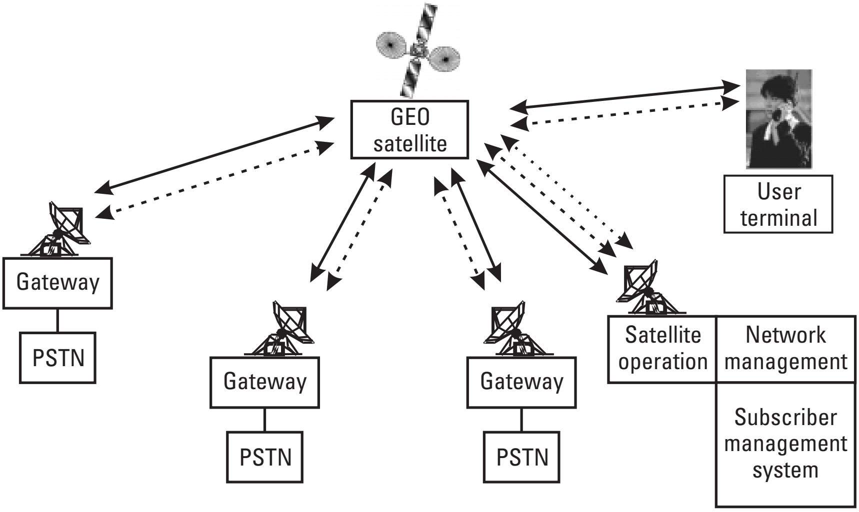


Figure 4.1 GEO mobile satellite and network architecture.

organizations that seek to standardize products and services and promote international business and social development. Reviewed below is a body of engineering practice that has developed from a terrestrial telecommunications foundation and has been adapted to the unique aspects of satellite communications ground segments [1].

At the same time that voice networks become more readily available and lower in cost to use, several important trends have appeared which are causing a paradigm shift in the engineering and operation of the global telephone infrastructure. As summarized in Table 4.3, these result in significant enhancement to the quality, affordability, and versatility of voice services from terrestrial networks. One of the more interesting innovations is

Table 4.3
Trends in the Use of and Requirements for Public Telephone Networks

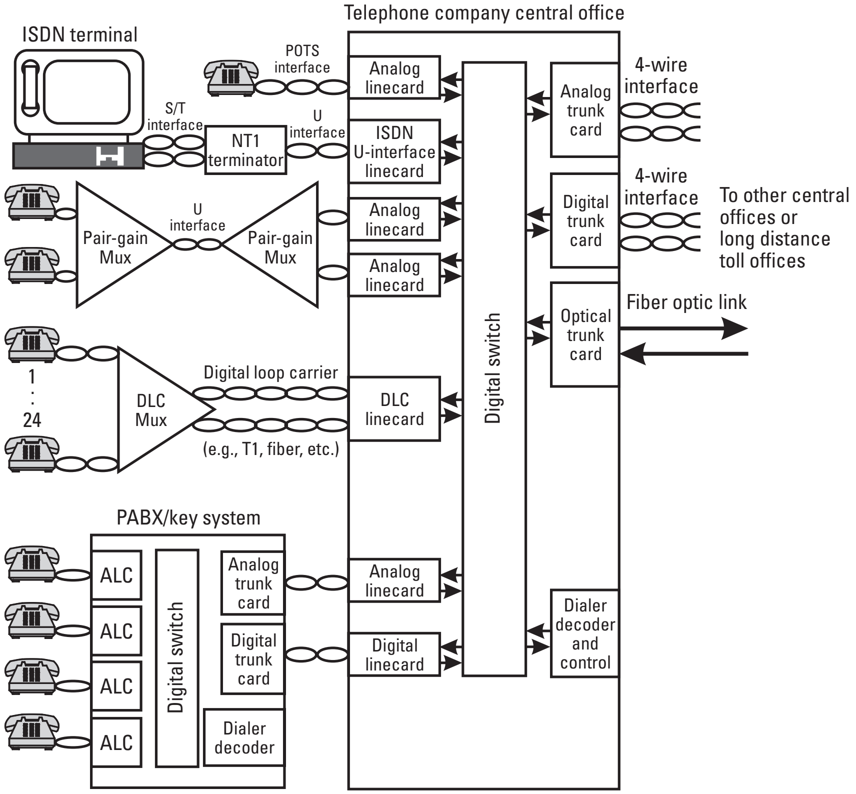
User Requirement	Approach	Impact
Voice telephony	Fundamental to the design and operation.	None (intrinsic to the current design and operation).
Fax transmission	Users employ Group 3 fax machines, with internal 9.6 to 14.4 Kbps modems.	Basically the same as voice telephony.
Advanced calling features	Digital telephone switches provide a range of features, like call waiting, voicemail, conferencing, blocking, caller ID, etc.	Available in developed countries when new telephone switches are introduced. Generally not available in developing regions trying to meet the basic needs of new subscribers.
Intelligent network features	Signaling Systems No. 7 (SS-7) supports online credit card calling, virtual private networks, integration of domestic with international networks, and improved network management.	Facilitates smooth interoperation of global telephone network; improves reliability of networks; allows greater usage by subscribers and reduced telephone service cost.
Internet access	E-mail and Web browsing extends connect times over access lines and local trunks; introduction of Voiceover IP (IPVoIP).	Local telephone service shifting to different models of service to allow extended connect times and greater bandwidth.
Impact of wireless networks	More people depending on wireless telephony for personal communications.	Reduced local calling for voice with possible implications for future network design.

Voiceover IP (VoIP), an evolving standard for adding telephone services to the Internet space. Satellite voice networks can still offer a relatively low-cost alternative for rural and developing regions until state-of-the-art terrestrial networks can be justified by the local economy.

The engineering and economic principles that underlie the design and operation of public telephone facilities have evolved over the 100 years that these services have been available. These are summarized in the following paragraphs as related to satellite networks that offer an alternative to terrestrial networks (or a complement, where appropriate). As the previous discussion indicates, the rules for network design are constantly being rewritten due to technology and economic/market factors. For example, in the United States the cost of long-distance service has dropped rapidly and subscribers can enjoy flat rates for nationwide calling of $.07 per minute or less, any time and to anywhere. Greater adoption of VoIP promises changing rates of this order on a global basis. At this level, subscribers need not even hesitate to make a call, day or night.

The public switched telephone network (PSTN) represents the common denominator for telecommunications throughout the world. To understand how satellite telephony services can interface with typical subscribers, we have to examine the architecture and interface arrangements of the PSTN. The basic interface arrangement for telephony may be split into the subscriber local loop and the trunks that connect between telephone-switching exchanges. Typical local loop configurations include two-wire analog local loop, four-wire analog loop to a private branch exchange (PBX), the digital loop utilizing the Integrated Services Digital Networks (ISDN) interface at 144 kbps, and digital loop carrier using time division multiplex (TDM) that provides either 24 (T1) or 30 (E1) digital voice channels.

Within the central office of the local telephone company we find the opposite end of the local loop connection, wherein a common digital format is used. Figure 4.2 provides a block diagram of a typical digital switching office used at the local level. This particular arrangement by Lucent Technologies is adaptable to all three general classes of exchanges: PBX, local exchange, and toll (long distance). Digital circuit switching operates on voice channels at 64 kbps, and offers digital trunks to interconnect with other exchanges. The majority of new installations in developed and developing countries employ digital fiber-optic transmission for these trunks, which assures high capacity and high quality for all interconnected services. In time, conventional time division switching with fixed bandwidth per channel will give way to dynamic packet routing using asynchronous transfer mode (ATM), but the timetable for this remains uncertain at the time of this writing.



Key:
ALC—Analog line card
POTS—Plain old telephone service
ISDN—Integrated Services Digital Network
DLC—Digital loop carrier
PABX—Private Automatic Branch Exchange

Figure 4.2 Interface arrangements for the PSTN. (Reference [2], with permission.)

An important feature to be provided along with the voice transmission is common channel signaling using Signaling System No. 7. This is a high-speed, complex data communication network architecture that supports the widest range of calling services and network management.

4.1.1 Bandwidth and Quality

Telephone instruments on the PSTN are predominantly analog two-wire devices. As shown in Figure 4.3, the network interface converts from two

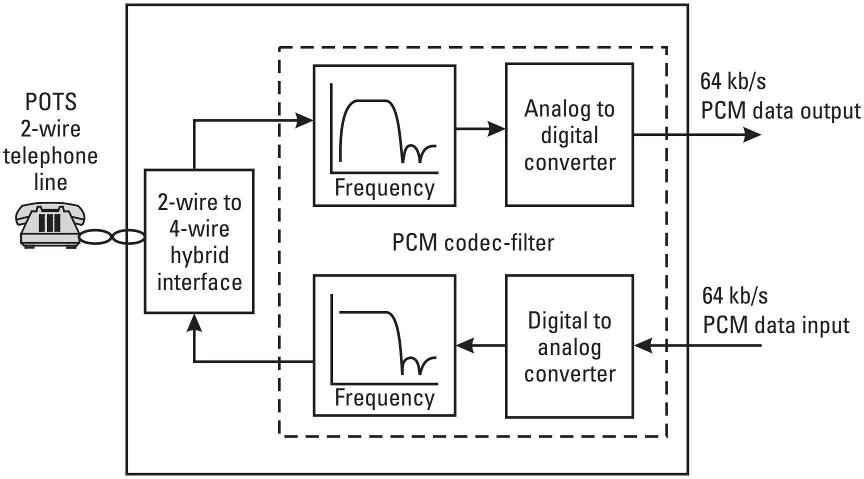


Figure 4.3 An analog telephone line-card that interfaces with the local loop on a two-wire basis and provides 64 Kbps PCM data input/output. (Reference [2], with permission.)

wires to four wires using a balanced transformer called a hybrid. As we will see later, the hybrid is the device most likely to produce echo impairment in long-distance telephone calls. The sending end of the circuit (incoming to the exchange) is filtered to restrict the channel bandwidth to the range 300 to 3,400 Hz. This allows the analog-to-digital (A/D) converter to sample at 8,000 samples per second, which is slightly more than twice the highest voice frequency. Using standard eight-bit-per-sample pulse code modulation (PCM), the output data rate is 64 kbps. (A technique called companding enhances the perceived signal-to-noise ratio, compensating for the imprecision inherent in A/D conversion.) On the receive side, the incoming 64 kbps bit stream is converted to analog and filtered to pass only the frequencies below 3,400 Hz. Receive analog information is inserted back into the two-wire line using the same hybrid.

Voice band channels are useful for relatively low data rate applications like fax, E-mail, and low-speed Internet access (i.e., less than 64 kbps). What makes this possible is the passband response and amplitude linearity of the PSTN, indicated by the attenuation mask in Figure 4.4. The dynamic properties of the channel for variable power levels, shown in Figure 4.5, allow analog telephone circuits to be very adaptable (which is an important reason they continue to be used extensively). The dynamic range is improved through companding, where Mu-law is used in North America and A-law in most other countries (particularly Europe). These characteristics are

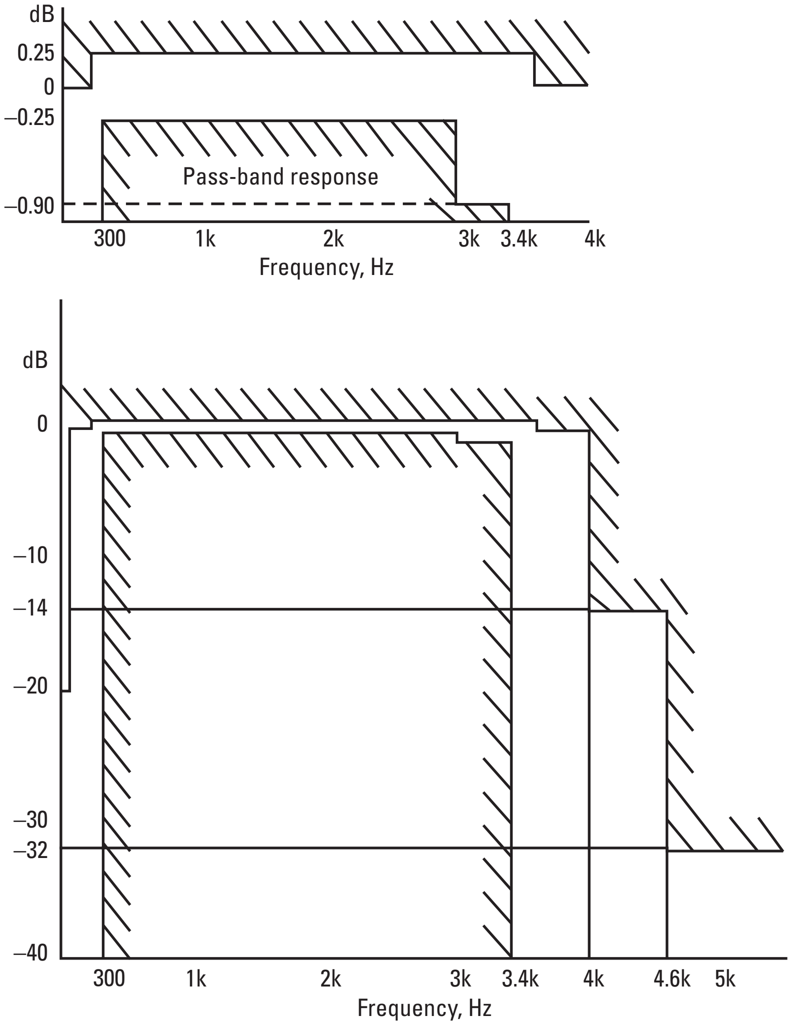


Figure 4.4 Typical frequency response for the input filter within the telephone line-card: passband and stopband. (Same attribute as Figure 4.3) Reference [2], with permission.

important to providing convenient data communication connections over the dial-up PSTN, owing in particular to the popularity of the V.90 standard for 56 Kbps modems.

Another important trend in the use of the analog local loop is the addition of high-speed access to the Internet through digital subscriber line/loop (DSL) technology. The family of formats and solutions indicated in

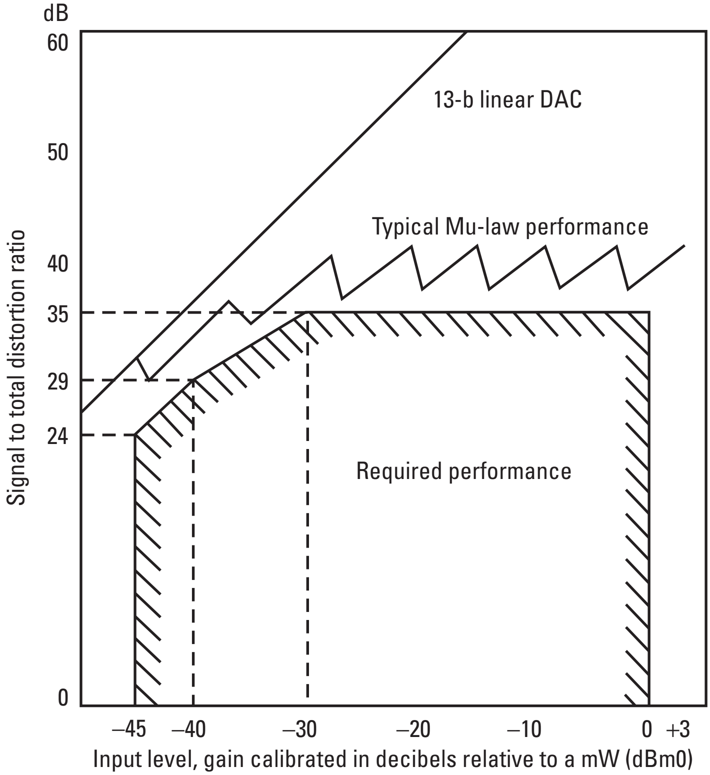


Figure 4.5 Signal-to-quantization distortion performance for Mu-law companded and 13-bit linear digital-to-analog conversion as compared with telephone network requirements. (Same attribute as Figure 4.3) Reference [2], with permission.

Table 4.4 defines most of the implementations that allow data rates in the hundreds to millions of bits per second to pass in both directions between the home and telephone central office typically using the existing copper twisted pair wire [3]. These break down into two categories: asymmetric DSL (ADSL), which provides a higher data rate for downstream (central office to home) than for upstream; and symmetric DSL (SDSL), which assures the same speed in both directions. Most implementations are going to ADSL to be consistent with the Web surfing tendencies of home users. Business users, on the other hand, may require exactly the opposite balance and hence SDSL or dedicated T1/E1 may be preferred. Since most forms of DSL share the bandwidth on a twisted pair of the voice channel, there is a distance restriction on the loop length. While these do not interface directly

Table 4.4
Implementations of DSL

Flavors	**Properties**	**Reference**
Asymmetric		
ADSL (full rate)	Offers differing upstream and downstream speed (up to 7 Mbps); allows high speed and voice to share the same line (analog and ISDN); requires special modem	ITU-T Recommendation G.992.1; ANSI Standard T1.413-1998
G.Lite ADSL	Medium bandwidth version for consumer market at speeds up to 1.5 Mbps downstream and 500 Kbps upstream	ITU-T Recommendation G922.2
Rate adaptive DSL (RADSL)	Nonstandard ADSL (standard ADSL also is adaptive)	
Very-high-rate DSL (VDSL)	Up to 26 Mbps on short local loops, within telephone plant; could be installed for VOD	
Symmetric		
Symmetric DSL (SDSL)	Being installed in U.S. by competitive local exchange carriers (CLECs); offers service comparable to LANs; popular in Europe	ETSI standard in preparation
High-data-rate DSL (HDSL)	Originates from 1980s; offers 1.5 or 2.3 Mbps without standard telephone service; replacement for T1/E1; uses two twisted pairs	ETSI and ITU standards in development
Second-generation HDSL (HDSL-2)	Replacement for T1 service (1.5 Mbps) without standard telephone service; uses one twisted pair	ANSI standard
Integrated services digital network DSL (IDSL)	Up to 144 Kbps using existing phone lines, through digital loop carrier, but provides "always-on" ISP service	

with telephone satellite networks, they represent data requirements that may need to be accommodated in some manner.

The digital standard for PSTN access at the subscriber level is ISDN, a service that provides 144 Kbps of active data subdivided into two 64-Kbps circuit-switched "bearer" channels plus one "data" channel (i.e., 2B + D). The standard provides for several interfaces, the selected one depending on the application and level within the network. Illustrated in Figure 4.6 is the reference model for the user interface, which provides for the following [4]:

- Terminal equipment type 1 (TE1), the user system consisting of an ISDN telephone instrument (which supports 144 kbps data and the 2B + D service), compliant fax machines and computer devices capable of using the ISDN service to establish 64 kbps connections via the 16 kbps D signaling channel.
- Terminal equipment type 2 (TE2), which are conventional consumer devices that cannot interface directly on a digital basis, including analog telephones, modems (e.g., V.90), and data terminal equipment (DTE) for other standards, notably X.25.
- Terminal adaptor (TA), an adaptation device to allow connection of non-ISDN devices that do not fall under the TE2 set of definitions. This is a catchall that requires a custom-tailored interface to perform such functions as A/D conversion, protocol adaptation, and channel signaling. An example is an ISDN custom PCI card that plugs into the PC. For a large enough user community, the particular TA device could become a de facto standard.
- Network termination 1 (NT1), the basic interface between the user terminal and the line to the central office. It supports the particular type of physical layer transmission medium being employed between the user and operator.

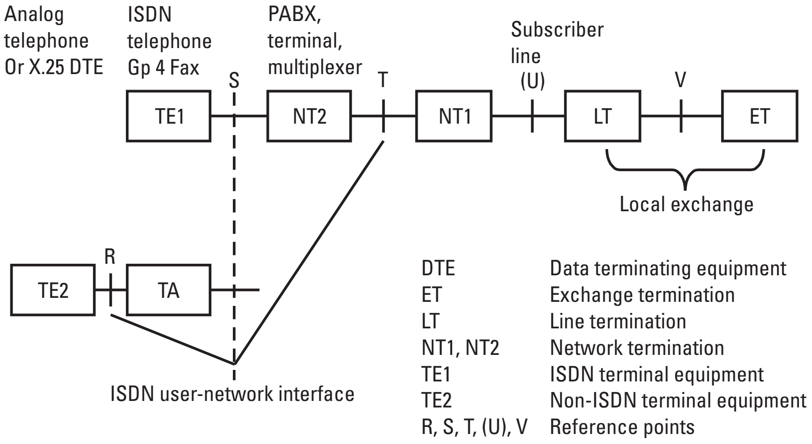


Figure 4.6 Standard user interface definitions for narrow-band ISDN.

- Line termination (LT), the other end of the link from the NT1, providing the interface to the exchange.
- Exchange termination (ET), the terrestrial interface to the ISDN-compatible telephone exchange, providing digital transmission at the bearer rate of 64 kbps and signaling via the packet switched network (using SS-7 within the PSTN).

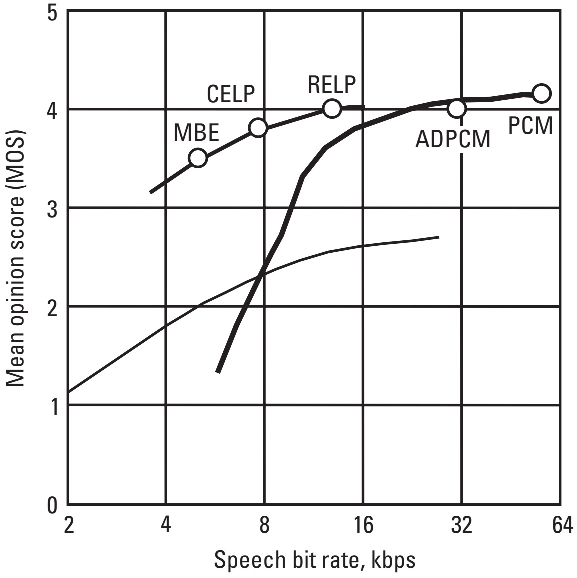
These interfaces are external to the ISDN switching fabric and digital trunks within the network itself. The ISDN standard was to have become pervasive in the digitized PSTN, but many telephone operators have been reluctant to include the necessary equipment in the telephone exchanges and local loop facilities. The advantage of 2B + D lies mainly with its ability to provide switched 64 kbps of clear-channel data. On the other hand, the wide availability of inexpensive V.90 modems and development of DSL have diminished the appeal of ISDN as it is currently defined. However, the infrastructure aspects of ISDN have proven successful around the world, particularly from a switching and network management standpoint.

The 64-kbps voice channel discussed previously delivers the highest-quality service for telephony. With the adoption of digital transmission from end to end, subscribers have become accustomed to clear-channel service without noise and with a minimum of disruption. Prior to digital fiber-optic and satellite services, networks were analog in nature and suffered from a variety of impairments. This is largely behind us; however, impaired voice communication has returned in the form of the digitally compressed speech used in mobile networks and via the Internet through VoIP.

The ITU has provided a process for evaluating the quality of voice channels that are impaired as compared to what subscribers expect. This is the mean opinion score (MOS) technique, embodied in the P series of recommendations [5]. A group of test users try out the subject telephone link, either on a listen-only basis or in an interactive two-way conversation. After completing the session, they rate the quality according to the following scale:

5. Excellent
4. Good
3. Fair
2. Poor
1. Bad

The scores are then averaged to produce the MOS for the particular run of the overall experiment. This rating system has proven quite reliable over the years and has maintained its popularity in the satellite communication and wireless industries. The MOS of a standard 64-kbps PCM telephone channel generally is slightly higher than 4.0, while significant impairment that reduces the naturalness or consistency of the channel will cause the rating to drop by one or more points. Ratings below 3.0 are generally not indicative of commercial quality. A typical trend curve is provided in Figure 4.7 [6]. Aside from bandwidth and naturalness, one of the most important impairments to a telephone conversation is delay, discussed in the next section.



Key to abbreviations:

PCM	pulse code modulation
ADPCM	adaptive differential PCM
RELP	residual excited linear predictive
CELP	code excited linear predictive
MBE	multi-band excited

▬ Waveform coders

▬ Hybrid coders

— Vocoders

Figure 4.7 Comparison of bit rate and MOS for various digital speech compression techniques.

4.1.2 Delay and Latency

The transmission of communication signals, whether via fiber-optic cable, copper, or radio waves, entails propagation delay. GEO satellites have long been used for telephone services, but it is often pointed out that the 260 ms of delay associated with one hop is a concern in this age of ubiquitous wireless networks and the Internet. One common principle for Iridium and Globalstar is to greatly reduce the end-to-end delay of a simple connection. In this section, we will consider how delay impacts voice services and what facilities are available to make satellite communication links suitable for commercial telephone services.

Delay can be broken down into the following categories:

- Call setup delay (also called post-dialing delay);
- Propagation delay;
- Speech processing (compression/decompression) delay;
- Switching and queuing delay (packet and circuit-switched);
- Delay-induced impairments (echo).

Each is reviewed in the following paragraphs.

Call setup delay is the time between when the dialing party completes entering the number of the called party and the called telephone instrument rings and is answered (assuming no additional delay for a slow response at the distant end). Since public telephony, whether using satellite or terrestrial access, requires that a connection be established before user information can be transferred, this is extremely important to the acceptability of the service. Figure 4.8 illustrates the steps involved and the following equation suggests how these contributors can be summed [7]:

$$T_{tot} = 3T_s + T_i + 3E(T_p) + E(W)$$

Where T_s is a constant (deterministic) delay assumed as for the time before a dial tone is delivered to the caller, the time for a ringing message to originate from the called end, and the time for the answer indication to be given to the caller, T_i is a constant time for a connection request to traverse the network to the called end, $E(T_p)$ is the expected value (statistical average) of the switch processing time at nodes A and B, and $E(W)$ is the expected value of the waiting time for the call request message to reach switch node B.

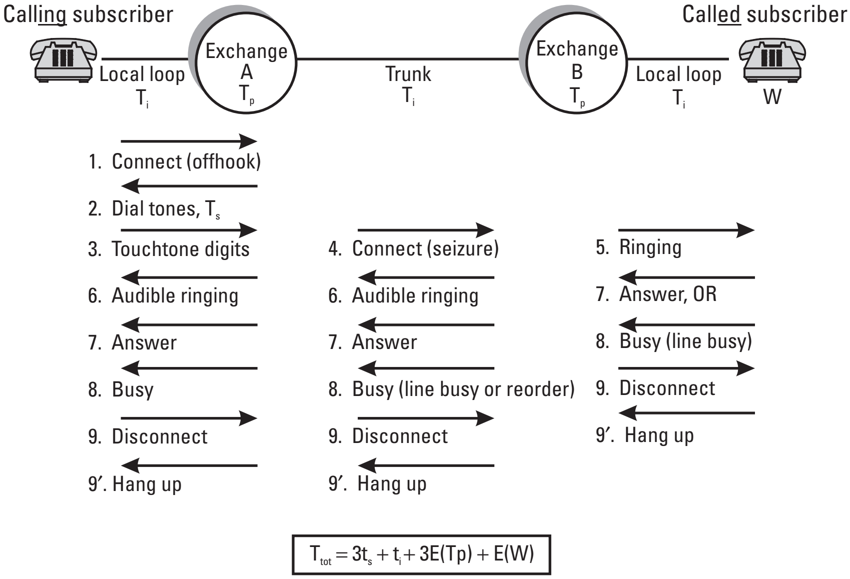


Figure 4.8 Telephone call signaling sequence, indicating transmission nodes, time duration, and delays.

This equation and Figure 4.8 are only meant to illustrate the components and how they combine to impact the call setup time. In a real case, the telephone engineer must evaluate in detail the delays at points of processing, switching, and transmission that affect call setup. This may entail actual measurements or, in the absence of a real system, computer simulation. Terrestrial fiber-optic networks that employ SS-7 produce low setup delay (less than 1 second), while the corresponding delay in cellular networks can be quite substantial (5 to 10 sec). Users who have experienced both have come to accept the difference in call setup delay performance.

Propagation delay is the time taken to travel over the cable or radio path. Presented in Chapter 3 and Figure 3.1, propagation delays for the LEO, MEO, and GEO altitudes are of the order of 7, 75, and 260 ms, respectively. LEO systems have the benefit of very brief single hop delay, which is comparable to delays experienced within telephone exchanges and multiplexing nodes. Multiple hops for LEO constellations would be quite acceptable, including intersatellite links. However, processing and switching onboard a satellite will add substantially to this delay, as discussed below. The propagation delay in MEO links is significantly greater, but is substantially below that of GEO satellites.

Propagation delay is only one component of the total budget, but for GEO systems it is usually dominant. Other contributors include compression and decompression processing and packet switching (if used). Experimental evaluation of MOS under various conditions of delay indicated that there is a cliff in the MOS rating curve at around 500 ms, a point where a significant fraction of telephone users find the circuit to be below 3.0. These scores are heavily influenced by the specific conditions of the call. For example, the best scores are always obtained for listening-only with steady-state signal quality. Two-way interactive conversation immediately brings out the delay characteristics and possibilities for speech clipping. However, subscribers involved in a mobile call with fading will tend to rate the same delay and quality with a higher MOS score (as much as half or a full point) because of a lower expectation.

Speech processing (compression/decompression) is a powerful technique to reduce the data rate necessary to transfer a voice channel. For heavy compression that reduces the bit rate to less than about 6 Kbps, the processing function will introduce a delay of its own. The amount of this delay, which is essentially constant and fixed for all telephone calls, results from the number of speech samples that are required and the actual processing time to perform compression and decompression.

Switching and queuing delay (packet and circuit-switched) is the consequence of internal node transfer and buffering of traffic within a network (satellite or terrestrial). While this is more of an issue for data communication in packet switched networks, it is still a natural consequence of having digital information transferred through time division switches. Also, as ATM is adopted throughout the global network infrastructure, processing and queuing delay will become a larger and larger factor in network design. An example of this type of consideration was already presented for call setup delay, but in this case, the delay affects all information flow after the connection is established.

Delay-induced impairment (echo) was a serious detriment for commercial satellite communications prior to the wide scale adoption of digital communications for international services. The problem was largely caused by a switching device called an echo suppressor, which was used in all long-distance circuits to block the reflection of echo. The half-second round-trip delay of GEO satellite links combined with poor echo suppression to produce service quality that was often unacceptable.

The solution to this problem was a digital device called the echo cancellor, which as the name suggests, actually removes the echo by canceling it out at the source. The hybrid at the distant end still returns the echo, but the echo cancellor introduces a negative-going replica of the echo to substantially remove it by numerical subtraction. When the circuit is first established, the echo cancellor goes through a training period to model the reflection path. The resulting attenuation of 40 dB is sufficient to make the circuit sound natural. In addition, hard echo suppression, which is digitally introduced, disables the path when only one person is actually speaking. The 40 dB of cancellation is really only required when both parties are speaking at the same time (called double-talking).

Echo cancellors are now standard in all long-distance circuits, whether on satellite or terrestrial links. This is important because all digital services involve the kinds of time delays previously discussed. For example, digital compression in terrestrial cellular networks introduces enough delay to demand that echo cancellation be provided within the digital cellular network (but not within the cellphone, since this is a four-wire telephone device). Likewise, the variable delay of the Internet requirement places an echo cancellation on VoIP services.

4.1.3 Traffic Engineering and Capacity

The design of traditional telephone networks was, for decades, viewed as a straightforward procedure using the analytical techniques derived by Agner Krarup Erlang (1878–1929), the first person to study this problem [8]. A Danish mathematician who worked for the Copenhagen Telephone Company, Erlang proved that Poisson's law of distribution governed the theoretical performance of circuit-switched networks. It is based on the principle that subscribers make calls independently of one another and these calls randomly arrive at the telephone exchange at an average arrival rate of λ calls per second. Mathematically, the probability of k arrivals in an interval of time T is given by:

$$p(k) = \frac{(\lambda T)^k e^{-\lambda T}}{k!}$$

This surprisingly simple formula tells us that the traffic demand can be specified exclusively by λ, the average arrival rate. What remains (to be discussed below) is how the switch and network respond to this demand. The individual demand per user is relatively small, perhaps only a few minutes during the busy hour. The traffic offered, A, to the exchange that routes calls over the channels is measured in Erlang (E), and defined as:

$$A = \text{call rate} \bullet \text{call duration}$$

For example, if the offered traffic is 0.5 calls per second and the average call duration is 30 seconds, then A = 15E.

The statistical property of telephone calls produces fluctuation in the loading of a particular trunk. If all N channels are in use, then the next call to arrive (the N + 1 call) will be blocked (e.g., the caller hears a "fast busy" tone). The telephone network could either put such calls in a queue awaiting available trunk capacity, or immediately reject the call (calls dropped). Since calls arrive randomly, N channels will carry fewer than N Erlangs of traffic, unless the network operator allows a very high frequency of blocked calls. Traffic loading depends on time of day, so we use the concept of the busy hour to form the basis of telephone network design.

A question might arise about the potential load offered by a single subscriber. Suppose that the typical subscriber makes one call during the busy hour and that this is two minutes in duration. Then the load for this subscriber is:

$$A = [1/3600] \bullet 120 = 0.0333 \text{ E or } 33.3 \text{ mE}$$

Thirty such users would produce the traffic load equivalent to one continuously occupied trunk.

A. K. Erlang derived the following mathematical relationship (the Erlang B equation for blocked calls dropped) to describe how this large community of potential callers will impact service on the trunks during the busy hour:

$$P_B = \frac{\frac{A^N}{N!}}{\sum_{n=0}^{N} \frac{A^n}{n!}}$$

where P_B is the probability that a call will be blocked because all trunks are busy.

For our simple example, 30 subscribers "offer" a traffic load of 1 E. If N = 3 channels in the trunk, this formula tells us that the probability is 6.25% (1 out of 16) that all channels are already occupied so that the next call is blocked, i.e., receives an "all trunks busy" signal from the switch. Normal telephone engineering practice would require P_B less than 1% (1 out of

100). A compliant design would provide five channels, for which $P_B = 0.3\%$. (It is impossible to have the precise solution at 1% because fractional channels are not physically realizable.)

The practical design approach in telephone network engineering is to predefine the value of P_B based on the quality of service that the network must provide to subscribers. In the PSTN, the accepted value is 1%, or one call blocked out of every 100 attempts. A graphical representation for a range of values of N and E is presented in Figure 4.9. We have chosen the Erlang B equation, above, which applies in the arbitrary case that all blocked calls are dropped, i.e., that the caller gives up if he or she cannot obtain the trunk for the call in question. The Erlang C equation is a refinement that allows for blocked calls to be held, that is, placed in a queue awaiting the next available trunk [9].

$$P_B = \frac{\dfrac{A^N N}{N!(N-1)}}{\displaystyle\sum_{n=0}^{N} \frac{A^n}{n!} + \frac{A^N N}{N!(N-1)}}$$

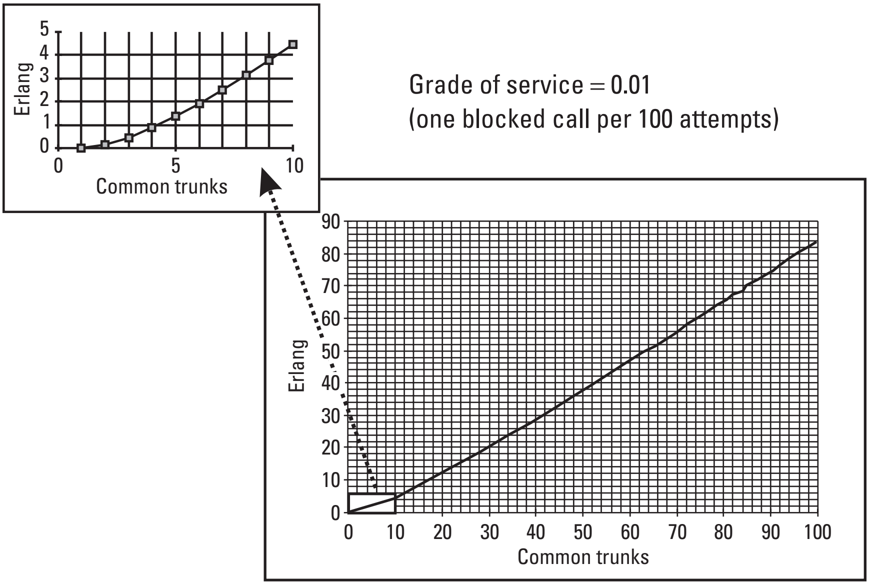


Figure 4.9 Traffic engineering design curves based on Erlang B equation (dropped calls cleared from network).

Whether blocked calls are cleared or held is largely a moot point, since neither of these elegant formulas is applied in practice. Rather, they provide a means to understand the statistical nature of telephone networks.

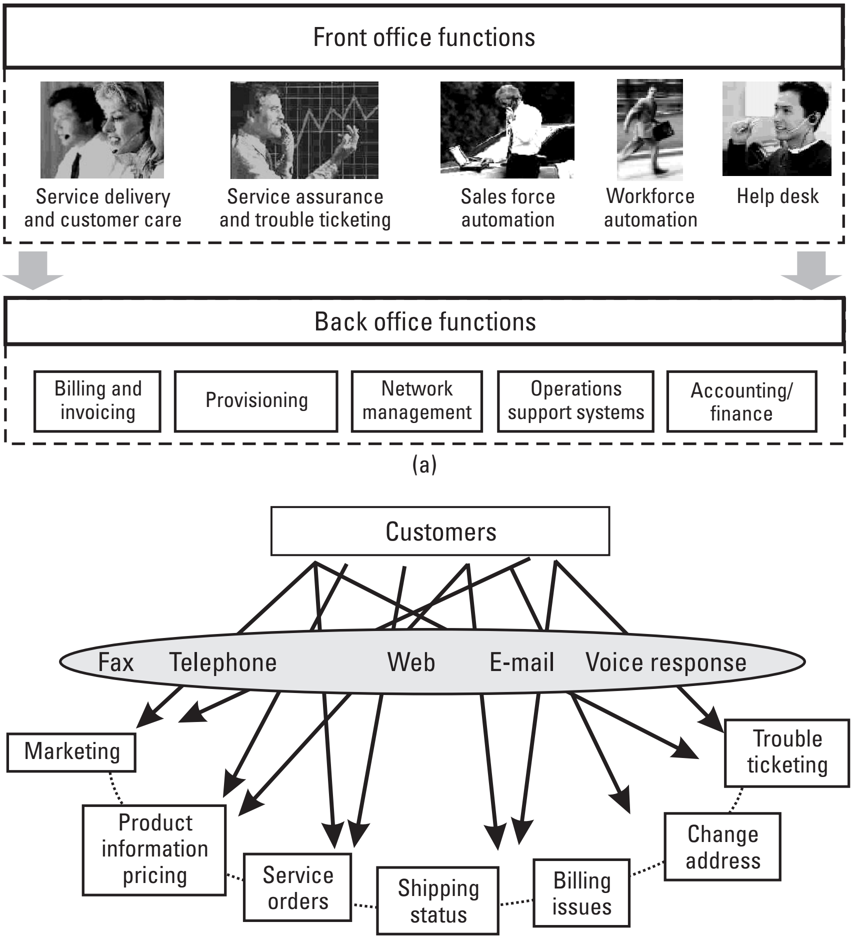
Estimating telephone traffic in a new network that has not been constructed, or worse yet for a service that does not yet even exist, is difficult, to say the least. The procedure seems to be to use either of A. K. Erlang's formulas as a rough starting point, and then to proceed with some kind of test case, possibly involving a pilot service on a testbed with live subjects. These data can then be used as an input to a computer simulation of part of the network under investigation. In this way, many scenarios can be tried and measured under a variety of conditions. The actual design to be installed should probably have excess trunk and switching capacity to alleviate blockages that can occur, and to reroute traffic as necessary during peak loading situations. It is hard to predict how new subscribers would react to a high level of blockage. For example, people in countries with low telephone density contend with such inadequate trunk facilities that one would speak of a call-completion rate as opposed to a call-blocking rate. More recently, first-generation cellular networks experienced overloading as their operators struggled to reach profitability during the early lean years. Subscribers on current fixed networks have become accustomed to having much better service levels available and so expectations are rising.

4.1.4 Service Management

Satellite telephone networks that deliver services to subscribers must be managed like a business, the most basic function being service management. One important aspect of service management is that the basic information about subscribers, their level of services, their usage records, and service bundles, must be generated within the network and made available to the operator of the network. The "hooks" for these functions must be introduced early in the development of ground segment components. Also, there could be requirements to pass such information back and forth across the interface to other service providers.

The basic arrangement of a service management system is illustrated in Figure 4.10 [10]. The following is a typical list of functions from the standpoint of telephone services that could be provided by a satellite communication ground segment:

- *User attributes and service bundle.* This identifies what services a particular user has requested and is entitled to employ. Along with this



(a)

(b)

Figure 4.10 Telecommunications-based service management systems: (a) information architecture for front-office and back-office management of the service business (illustration courtesy of Clarify, Inc.); (b) customer interaction in the New Economy.

goes a pricing arrangement (referred to as rating), including the basic rate for service and usage-based pricing (e.g., price-per-minute during day and evening).

- *Accounting*. Accounting management deals with the generation and processing of information to measure usage of network resources (ground and possible space as well) and creates the call detail information fundamental to creating call records [11]. This is the basis of billing, discussed below.

- *Billing.* This uses information gathered as the user places calls, taking account of user attributes and the associated service bundle. Call records are gathered from the network and stored for monthly resolution. The bills can be printed and mailed, although there is greater interest in making billing data available online over the same network or the Internet. Payments must be collected and billing resolved so that the proper steps are taken to reduce delinquency and fraud. Some of the steps that can be used to improve payment performance are to employ prepaid calling cards or smart cards or to have all bills cleared with credit cards.
- *Mobility management.* Satellite-based networks typically allow users to move anywhere within the coverage area. Therefore, there should be a scheme for authenticating a user who activates his or her terminal from an arbitrary location and attempts to place calls. The classical scheme for doing this is the visitor location register (VLR) common in the GSM digital cellular standard [12].
- *Service provisioning and activation.* The service provider or reseller who acquires subscribers and puts them into the network in the first place needs a user-friendly system to activate the service on an individual basis. For a fixed-line solution, this also involves sending an installer out to the customer's location. Satellite services can permit customers to provision their own access by purchasing the necessary terminal from a distributor and subsequently doing a self-install at the business premise or home. When activated, the network should become more or less transparent to the subscriber, with he or she only having to make infrequent service inquiries regarding problems or new features. This function interfaces with the activation function described previously.
- *Fraud management.* Wireless networks have become the targets of fraud, the unauthorized use of the network without paying. The two paths where fraud is encountered are at the user terminal level, e.g., using a legitimate device available through the network to place illegitimate calls, or accessing the network over the radio link with a specially created or modified user device that intentionally bypasses the call accounting and billing systems. The former occurs when a user simply stops paying the bills when they arrive; this can be

> resolved by following up quickly on non-payers. The second abuse is usually more difficult to reduce because the perpetrators probably have inside information on the weaknesses of the system. Reducing this problem means either locating these criminals or blocking their access by closing the loopholes. Fraud management is fundamentally a game of cat and mouse, in which the network operator (the cat) must be extremely vigilant and aggressive to uncover the abuses and cut them off; the wrongdoer (the mouse) continues to find new and more intriguing ways around the system. Other security concerns include integrity of the data discussed in this section and limitation of damage due to violations of security processes.

The previous discussion was only meant as an introduction to the general topic of service management and should not be viewed as a comprehensive list of requirements for a given telephone ground segment. The requirements can only be determined once the specific nature of the services and network design has been determined.

The various computers, software systems, information networks, and supporting organizations that accomplish the above objectives represent a substantial investment and require considerable ongoing maintenance. In recent years, these functions have migrated to specialized software suites developed under the broad panoply of *customer relationship management* (CRM). In the telecommunications industry, the leader in providing comprehensive and well-integrated CRM solutions is Clarify, now part of Nortel Networks. One discovers quickly, however, that while CRM is very attractive for the operator-customer interface (including the provisioning aspects), it does not as yet address recording of calls and billing. These functions are complex in their own right and have, in the past, been addressed by switch manufacturers like Alcatel, Lucent, and Nortel. However, specialized billing software, with special features for flexible rating (e.g., complex billing algorithms to lure customers), have appeared on the market. One of the companies that provides billing solutions is Geneva, based in the United Kingdom. In the future, it is possible that, with the complexity of these software systems, operators will choose to outsource CRM and billing to major computer service firms like EDS and IBM.

4.1.5 Synchronization with Terrestrial Networks

We have already reviewed many aspects of the service and interface requirements of telephony, whether via an earth station or terrestrial PSTN facility.

Beyond these basic requirements, there are a host of other issues that can only be addressed once the local situation is understood and the appropriate scenarios described in detail. This requires a lot of research on the ground, including a thorough investigation of the existing PSTN standards and facilities. The fact that a given country or region has adopted an ITU, ETSI, or Bell System standard does not guarantee that new earth stations or switching equipment will perform properly.

One of the questions that often gets overlooked is that of synchronization of timing and clocks. Two fundamental approaches are applied.

- In plesiochronous mode, timing is inherently flexible to allow for a wide disparity in local timing. This is made possible by including dummy bits that can be removed or inserted at each node.
- Synchronous mode relies on close alignment of timing throughout the network. This demands high accuracy of individual clocks, of the order of 10^{-11}, in order to achieve maximum throughput without dummy bits. The benefit is provision of end-to-end clear channels for precise data and applications that demand this type of performance [13].

The goal of synchronous mode is to provide a stable reference that is traceable to a highly accurate primary reference source (PRS), such as a rubidium atomic clock, the U.S. National Institute of Standards and Technology (NIST) WWV radio broadcast, GPS, or the like. This timing can be further distributed over the satellite to all locations, a technique that can be successful as long as the local environment at the earth station is compatible. For example, the satellite network may be properly tied to the PRS, but the local switching office (which is independent of the satellite network) may not be so exact.

The slippage of clock timing can cause unexpected difficulties in the operation of certain types of equipment. Voice connections may not even notice a clock slip, which might cause nothing more than a click or dropout of a few milliseconds. Connections involving voice-band modems will not fare as well, possibly causing a session to end abruptly. A slip in a fax transmission could produce a vertical gap of a few millimeters. Encrypted services such as voice or data are very susceptible to slips, as these interrupt synchronization, a problem that impacts digitally compressed video as well. The solution may require forcing the newly introduced satellite network to follow the local clock, or vice versa.

4.1.6 Signaling in Telephone Networks

A second important consideration in the telephone interface has to do with the signaling system used for call setup, processing, and billing [14]. Telephone signaling has developed slowly over the decades because of the enormous established base of telephone exchanges in each country. The two basic types of signaling are supervisory and interregister (i.e., numerical). Supervisory signaling directs the process of initiation and teardown of the call. On the originating end, a signal is transmitted to "seize" the line, that is, to inform the other end that there is a demand for service on a particular line. The other end acknowledges the seize and holds the line open for the subsequent interregister signaling, which conveys the city or area code and local number. Table 4.5 provides a summary of techniques for each aspect of signaling on subscriber loops and trunks [15].

On most telephone networks, supervisory signals are energized through a separate pair of wires on transmission equipment, called the "ear" (E) and "mouth" (M) leads. The E lead accepts signaling input from the originating switch or transmission line, and the M lead transmits signaling to the next link in the chain or the terminating switch. With conventional transmission equipment, supervisory signaling is carried at a frequency of 3,825 Hz (3,700 Hz in North America), which is outside of the voice frequency band ("out of band") yet follows the same path as the associated voice channel. Subscribers cannot hear the tone because bandwidth filtering for

Table 4.5
Comparison of Loop and Trunk Signaling in Telephone Networks

Type of Signaling	Technique	Standard
Subscriber loop		
On-hook/off-hook	DC current	—
Dialing	Multifrequency code	—
Caller ID	PSK ~1,500 bps during ringing	—
Trunk, analog		
4-wire, supervisory	In-band	R1
6-wire, supervisory	Out-of-band, E&M	R2
Interregister	Multifrequency code	R1, R2
Common-channel signaling	Out-of-band, packet switched network, 64 Kbps	SS-7

300–3,400 Hz in the transmission system blocks it. Some older North American transmission systems use "in-band" supervisory signaling, also called single frequency (SF) signaling, with an audible tone frequency of 2.6 kHz. In this case, a narrow "notch" filter is used to remove the tone from the voice frequency bandwidth. AT&T discovered that telephone pirates, called "phone freaks," made free calls by using "blue boxes" that generate SF signals and interregister signals (discussed below) to fool long-distance switches into allowing nonpaying calls out of band through the network. Phone freaks are largely out of business by virtue of common channel signaling (SS-6 and SS-7), which do not recognize the tones coming from blue boxes.

During the 1960s and 1970s, the most common form of interregister signaling used the voice channel itself to pass the dialing digits, also called address information. In the in-band approach, each digit (0 through 12 or 0 through 16, depending on the system) is coded into a pair of audible tones, there being seven or eight different tones to be employed. You can listen to the combinations when you touch the keypad of a telephone. After the digits are received and interpreted by the switch at the other end, the call is completed to the distant party. When the party answers, the distant switch returns an appropriate supervisory signal to confirm the connection. Now the switching equipment can track the duration of the call for billing purposes. The final step occurs when one party hangs up, causing a "clear" supervisory signal to be transmitted. The switches take down the circuit connection so that another call can be connected through. Billing information is also recorded.

The predominant in-band signaling systems in North America and Europe are called R1 and R2, respectively, and were adopted as international standards by the ITU-T. The R2 system is also used in many countries of Asia and Africa. Variants of these systems have been developed to make them compatible with satellite links to bypass the "compelled" mode, which holds the tone on until a response has been received from the distant register. If a satellite link were introduced, this handshaking would add unacceptable time delays to set up a call. Semicompelled signaling was devised wherein tone pulses of fixed duration are used. The delay for call setup is reduced because the sending equipment waits only once per digit for acknowledgment before sending the next tone pulse.

International trunks over transoceanic cables impose additional complexity to the problem of interconnecting telephone exchanges. A form of semicompelled tone signaling, called ITU-T System Number 5 (SS-5), resulted from a joint development effort of AT&T Bell Laboratories and the

British Post Office. This system operates effectively over transoceanic cables that are equipped with time-assignment speech interpolation (TASI) equipment, and over satellite circuits as well. TASI is a circuit-multiplying technique that detects an active speaker at the input to the cable and switches the speech into an available channel. In digital telephone trunking, TASI is replaced by digital circuit multiplication equipment (DCME), which combines an advanced form of speech interpolation with voice compression through adaptive differential pulse code modulation (ADPCM). This can yield an effective gain in voice channels of from four to eight, depending on the number of voice channels available.

AT&T first introduced common channel interexchange signaling (CCIS) to combine the supervisory and interregister functions on a dedicated out-of-band data network backbone. Later, CCIS was adopted by the ITU-T as Signaling System Number 6 (SS-6), the precursor to the more advanced version called SS-7. Under this concept, the supervisory and interregister signaling information is stripped from the telephone circuit at the originating exchange and transmitted as a message over an entirely separate data communication link that runs in parallel with the voice network. The message format takes advantage of the principles of packet switching and processing, which can be made extremely reliable. Error-detection techniques such as cyclic redundancy checks and forward error correction allow the data communication network to yield messaging reliabilities of greater than 99.9999999%. The developers of SS-6 felt that this type of reliability was necessary because the distant exchange would not be in direct contact with the originating exchange when completing the call. Also, SS-6 does not interfere with the operation of DCME and is not seriously affected by satellite propagation delay.

The common channel signaling approach is integral to SS-7, whose datagram message structure is based on the Open Systems Interconnection model and is also the signaling system for ISDN. The physical channels transmit synchronous data at a rate of 64 kbps, raising the number of telephone channels within a given group by several multiples. At the link level, SS-7 employs high-level datalink control (HDLC), a robust protocol employing the “look back N” technique. The higher level protocols of SS-7 achieve reliable end-to-end message routing and facilitate the specific content structure of call processing messages. SS-7 includes ample capacity for new features for use by subscribers and by the telephone operators themselves, but this capacity must be activated by all players for the potential to be realized.

4.2 VSAT Data Networks

Very small aperture terminal (VSAT) satellite communication networks were introduced during the early 1980s in response to pent-up demand for a low-cost and reliable alternative to leased analog or digital telephone lines for business data communications. Consisting mostly of user-premises, small earth stations and a centralized hub main earth station, the VSAT network gained popularity by appealing to a segment of U.S. corporations, offering more effective service than that provided by analog telephone lines employing the 9,600 bps modems of the day. The first users in the United States included Wal-Mart, then an up-and-coming retailer that built its stores near small towns in rural America; Chevron, the first oil company to place VSATs at each of its company-owned filling stations; and Toyota of America, which wanted to enhance the quality of service provided to the first buyers of their new Lexus automobiles.

In reality, VSATs can be configured for essentially any kind of telecommunications service (not just data, but also telephone and video as well) and are suitable for installation on customer premises. The most visible part of the VSAT is its antenna, which typically is less than 3m in diameter. There is a trend to push size down to approximately 1m in an effort to reduce the cost and difficulty of installation on typical buildings. Ultimately, we wish to be able to locate VSATs at homes to encourage personal use and small office/home office (SOHO) applications. At the time of this writing, this type of equipment was not on the market, with the possible exception of receive-only data communication terminals like those provided for DirecPC and Cyberstar. The direction that we are going in is to allow two-way communication over the satellite, eliminating any terrestrial connection (unless the user wishes it for backup communication and to provide another type of application such as video on demand).

In the following sections, we provide a brief review of the basic arrangements that VSATs offer, as well as the types of applications that are currently supported. This should be viewed as the starting point for developing a specific ground segment and for defining requirements for the earth stations themselves.

4.2.1 VSAT Network Topologies

VSAT networks fall into two broad categories: the hub-based "star" VSAT network, which provides a star type of topology, and the "mesh" network, which allows connections between any pair of VSATs. Acting as the common relay point, the space segment should be fairly transparent to the operation of the network, unless we are speaking of a processor-based repeater design. The latter may offer hub characteristics without the overhead and added delay of a ground-based hub.

As with any satellite network, VSATs must employ a multiple access (MA) system that is both reliable and efficient. These methods include TDMA, which is the most common for VSAT networks, CDMA, which is gaining popularity in mobile networks, and FDMA, the original MA technique and still quite useful for mesh topologies. Hybrid networks that adopt two of these techniques have also appeared as a way to address some of the dynamic aspects of VSAT networks. For the mesh, all earth stations must transmit with sufficient power to reach any other earth station in the network (to permit "any to any" types of connections), while in the star network there can be an imbalance of power favoring the VSATs. This recognizes that very large networks of small earth stations (with antennas of around 1m and SSPA output power of less than 5 watts) benefit from using a common hub with a large antenna (6m to 10m) to be capable of receiving the low-power VSAT transmissions. This assumes the common analog bent-pipe type of repeater with linear (or near linear) gain characteristics. Later, we provide link budget examples to emphasize this particular point.

In the mesh network, we must still provide some form of centralized network control and management, since much of the communication will benefit from a demand assignment (DA) scheme of some type. This central control could use one of the mesh network earth stations to access the space segment. The purpose of allowing this is to permit any station to request a connection to any other and to obtain the necessary space segment bandwidth and power for the duration of this connection. The central control function is vital to the proper operation of the ground segment, which cannot function without the ability to assign and remove the necessary links and manage the network in a service sense. Channel capacity to be assigned can be in terms of therefore bandwidth within the transponder or baseband bandwidth in packet form.

The demand for VSAT applications has pushed the operating frequency higher into the Ku and Ka bands, where spectrum is more readily available than it is at C, S, or L bands. This means that VSAT links must accommodate significant rain attenuation, particularly in temperate and tropical regions of the planet. As discussed in Chapter 2, it is possible to obtain good link availability (i.e., at the 99.9% level) provided that there is adequate link margin. Another aspect of using Ku and Ka bands is that terrestrial interference from microwave stations will not be present in general and so VSATs can be operated from almost anywhere as long as the satellite (or satellites) can be seen without ground obstruction.

4.2.2 Computer Network Requirements

The first step in understanding where VSAT mesh and hub networks fit in is to examine how companies apply data communications to solve business problems. We include under "business problems" the kinds of applications used by nonbusinesses, including governments, educational institutions, and nonprofit organizations. All of us, then, can acquire the computer hardware and software resources to automate routine functions, access and process data, publish and distribute information, and more recently, engage in electronic commerce at the retail and interorganizational level. Some general examples defining the uses of VSATS are:

- As a replacement for an existing star or mesh network that serves an organization's data communication needs (this existing network may also provide other forms of communications, notably voice and video);
- To activate a new strategic information technology (IT) application to achieve a business goal of becoming more competitive in a particular market;
- As an integrator of wideband services, where the organization had been using a number of different networks and services, each addressing a different requirement;
- For rapid installation or expansion of an existing network, taking advantage of the unique features of satellite communication;
- As a platform for a corporate Intranet that addresses a wide variety of locations within a country or region, or even on a global basis. An Intranet is a closed computer network that uses the open architecture of the Internet, the Transmission Control Protocol/Internet Protocol (TCP/IP), but limits access and applications to those that support the internal needs of the organization.

Computer networks have long been viewed in terms of a layered architecture, typically following the seven layers of the Open Systems Interconnection (OSI) model [16]. Modern data communication theory and practice is literally built upon the concept of protocol layering, where the most basic transmission requirement is at the bottom, and more complex and sophisticated features are added one on top of another. As we move up the "stack," each layer obtains a service form from the layer immediately below. In this way, the details within the layer can be optimized for performance and isolated from the other layers. At the very top of the structure is the actual information processing application that is required in order for the network to carry out its function. These principles are summarized below with a definition of the specific layers (which are normally arranged with Layer 1 on the bottom, but listed here in ascending order, for clarity).

- *Layer 1—Physical.* Provides the mechanism for transmitting raw bits over the communication medium (cable, wireless, or satellite). It specifies the functional, electrical, and procedural characteristics such as signal timing, voltage levels, connector type, and use of pins. The familiar RS-232 connector definition is a good example of the physical layer. On satellite links, the physical layer is built into the modem and MA elements and must be developed to maintain a reliable point-to-point or point-to-multipoint transfer of data in the presence of link fades and noise-induced errors and interference.
- *Layer 2—Data Link.* Provides for the transfer of data between adjacent nodes or connection points, either by a dedicated point-to-point line (e.g., a T1 private line or a satellite duplex link) or a medium capable of shared bandwidth (e.g., an Ethernet cable or satellite TDMA channel). VSAT terminals and hubs must incorporate a basic protocol for maintaining reliable data transfer, often including a scheme for automatic retransmission of lost packets, done in a manner that is invisible to the higher-layer protocols by the computer application.
- *Layer 3—Network.* Responsible for routing information from end to end within the network, allowing for multiple data link paths. This may involve decisions as to the most effective route through the point-to-point links that comprise the network. A VSAT network may serve as one of these links and hence would have to interface properly with the network layer. Popular examples of the network layer are the X.25 protocol used internationally and the Internet Protocol (IP) that is employed in the majority of router-based private networks, as well as in the Internet itself.
- *Layer 4—Transport.* Provides another level of assurance that the information will properly traverse the network, from end user to end

user. Two transport layer services are commonly available: *connectionless*, which transfers packets of data, one at a time; and *connection-oriented*, where a virtual circuit is first established before sending multiple packets that make up the entire conversation. The familiar TCP layer of TCP/IP provides a connection-oriented service to computer applications. Generally, the satellite system tries to use the terrestrial transport layer, although some form of spoofing (discussed later) may still be needed to improve throughput and reduce the average delay from end to end.

- *Layer 5—Session.* Somewhat more complicated than Layers 3 and 4, but provided to instill yet greater degrees of reliability and convenience of interface to applications. It manages the data exchange between computer systems in an orderly fashion to provide full duplex or half-duplex conversations. One important service is that of reestablishing the connection in the event that the transport layer is interrupted for some reason.
- *Layer 6—Presentation.* Provides syntactic and semantic services to the application layer above. What this means is that the presentation layer is inserted to resolve the complexities between transport/network layers and the more simplistic needs of the actual application that employs the network in the first place. Examples of syntax include encryption, compression, and numeric encoding (e.g., ASCII).
- *Layer 7—Application.* Includes the actual data communication applications that are common in open systems, such as file transfer, virtual terminal, electronic mail, and remote database access. We refer to these as applications because they include not only the protocol elements that support specific types of information but also features and facilities that ultimately interact with the end user. Most nonexpert users will not use the application layer directly, instead relying on specialized software within the computer to improve the interface and functionality. For example, most Internet surfers use the E-mail package supplied with the browser. This package, in turn, will engage Layer 7 electronic mail services (e.g., SMTP) to do the actual function of addressing, sending, and receiving message traffic.

The previous discussion was, of necessity, somewhat abstract as it relates to the logic of how networks are put together and interoperate.

Table 4.6 offers examples from terrestrial and satellite worlds to provide more of a practical feel for these aspects.

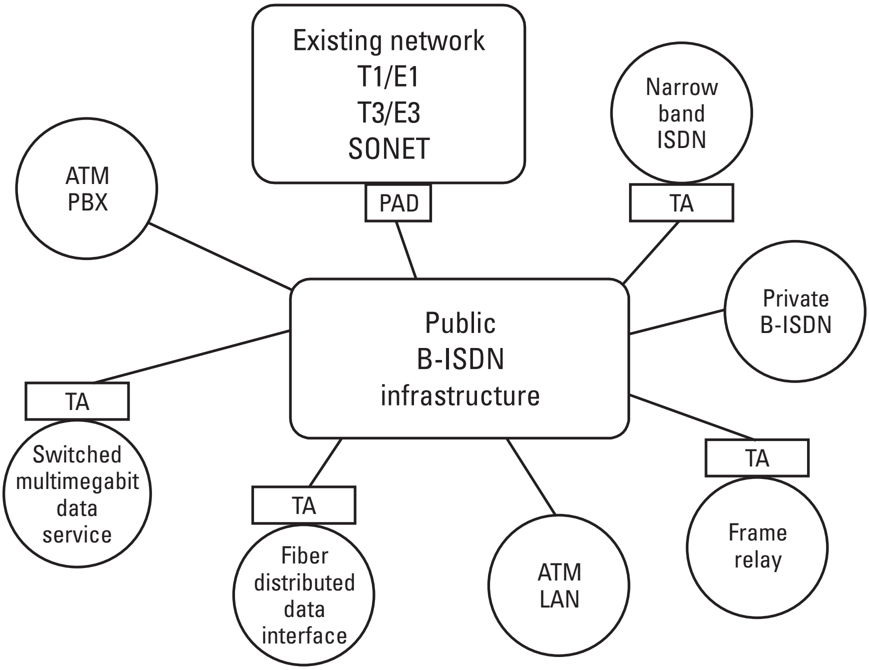
We see from Table 4.6 that satellite links differ significantly from terrestrial data communications at the lowest three layers: Physical, Link, and Network. This takes account of the substantially different technical environment of satellite links, particularly the delay, connectivity, and error-production mechanisms. Terrestrial links have their foibles regarding errors picked up along the route and delays encountered from links, nodes, and contention for these resources. The key point is that the widely used protocols, such as TCP/IP, X.25, and Frame Relay were designed for this environment. They can accommodate satellites as well, but usually in a nonoptimum way. What we find is that the protocol must be adapted to satellites or performance will probably suffer. An effort under the panoply of the Internet to create a facility called Multi-potocol Label Switching (MPLS) offers a long term solution for TCP/IP.

ATM is gaining in popularity for a wide range of applications and can support IP traffic as a backbone network. This allows integration of multiple information forms onto a single network architecture and achieves a higher degree of information transmission efficiency. An example of such an architecture, which had in years past been referred to as broadband ISDN (B-ISDN), is presented in Figure 4.11. In it, we see that a variety of user devices and access systems can connect to the ATM backbone using either a direct connection or a terminal adapter (TA) that converts the user protocol to ATM. Most popular at the time of this writing were TAs for LAN

Table 4.6

Protocols of the OSI Stack Layers, for Terrestrial and Satellite Implementations

Layer Definition	Terrestrial Examples	Satellite Examples
1. Physical layer	Fiber-optic cable, RS-232 connection, GSM cellular	QPSK modem-to-modem connection
2. Link layer	Ethernet, SONET, SDH	TDMA, ALOHA
3. Network layer	X.25, ATM, IP, Frame Relay	Proprietary protocols
4. Transport layer	TCP, SPX/IPX	Proprietary protocols
5. Session layer	Sockets, NetBIOS	Same as terrestrial
6. Presentation layer	ASN1	Same as terrestrial
7. Application layer	http, SMTP, X.400	Same as terrestrial



TA—terminal adaptor
PAD—packet assembler/disassembler

Figure 4.11 Application of existing and evolving networks to a public B-ISDN infrastructure.

protocols (e.g., Ethernet and Token Ring), IP (usually incorporated into the router), and Frame Relay. It should be kept in mind that whatever protocol is used on the input TA must also be used on the output TA (e.g., an Ethernet TA on one end must be matched by an Ethernet TA on the other).

The protocol layering of ATM is illustrated in Figure 4.12, indicating the ATM layer (with its 53-byte packet) at the middle. This amounts to a basic network layer that supports all types of services, such as constant bit rate (CBR) connections for voice and conventional video, and variable bit rate (VBR) for dynamic data communication involving IP and similar protocols on the end-user side. It is the function of the ATM adaptation layer (AAL), which resides just above the ATM layer, to perform the packet assembly and disassembly (PAD) functions that these user services require. At the physical layer we see a convergence sublayer to insert the 53-byte

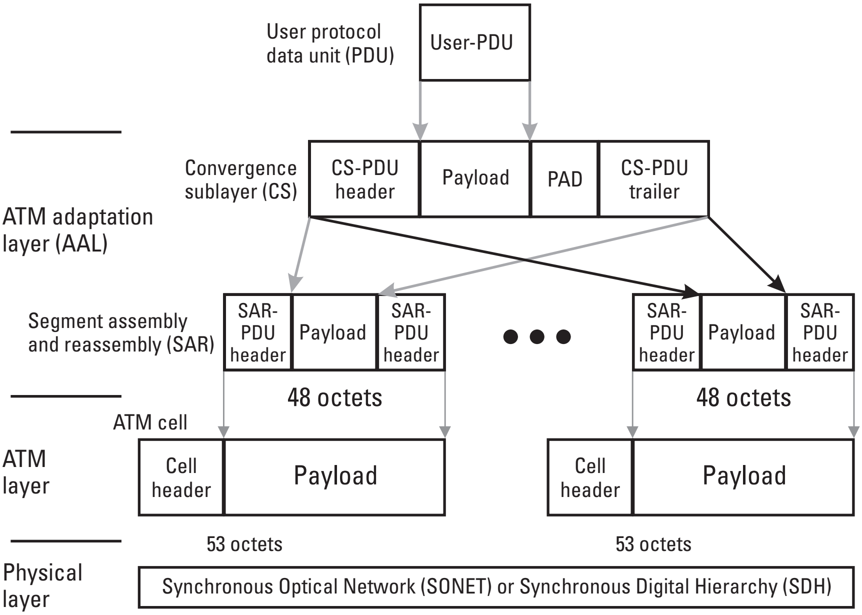


Figure 4.12 The protocol layers and cell structure of ATM.

asynchronous cells into the digital transmission scheme, which is to carry ATM-based data from point to point. Examples of convergence sublayer implementations include SONET/SDH, T1/E1, and Ethernet.

While ATM is growing rapidly at the time of this writing, its use over satellite links is limited—mostly to point-to-point connections between countries over INTELSAT, PanAmSat, or another international or regional satellite system. However, ATM plays a role in broadband satellite networks, particularly with digital processing repeaters. Fixed-length cell transmission will integrate the ground segment with telecommunication networks as public ATM gains acceptance. Readers should keep in mind that the 53-byte approach of ATM is rather simple; the complexity comes with the adaptation of ATM to the wide variety of higher-layer protocols and applications that exist and will appear in coming years. Considerable information on the evolution of ATM, including current and pending standards, can be obtained from the ATM Forum (www.atmforum.com), an industry-supported body that assists both suppliers and users of ATM technology.

Another commonly used protocol and network architecture is Frame Relay, which is based on the Link Access Protocol—Balanced (LAP-B) international standard. Frame Relay was introduced in the early 1990s to exploit the low-cost bandwidth made available on the newly developed fiber-optic networks in North America and Europe. Major IT vendors like IBM, AT&T, and Cisco Systems made the hardware and software available to facilitate introduction of Frame Relay systems and services on a national and international basis. In addition, the ATM Forum promoted implementation of Frame Relay over ATM to provide for the interoperability and longevity of both of these standards.

Figure 4.13 illustrates a typical Frame Relay network supporting a business with several major locations. The advantage of this approach is that the user pays for the access speed needed at each location (they need not be equal but rather are tailored to local requirements) and for the total aggregate data throughput of the network (termed the *committed information rate*, or CIR). This reduces costs substantially as compared with using dedicated leased lines between these locations. Additional information on Frame Relay can be obtained by consulting the Frame Relay Forum (www.frforum.com).

Throughout the discussion of efficient networks using ATM and Frame Relay, the basic premise remains that IP will dominate data applications. Much has been written about the design, features, and application of IP [17]. With its variable-length packets, ability to offer networking solutions for a wide range of protocols, and support through innovative

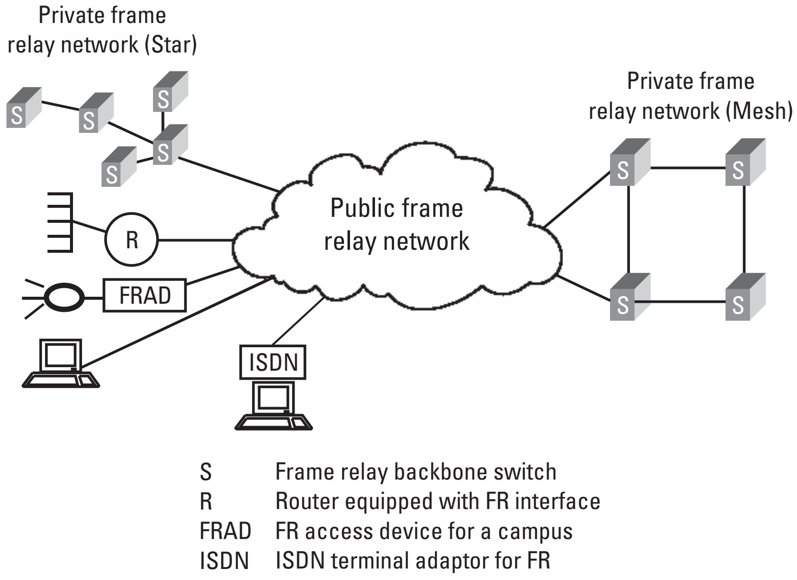


Figure 4.13 Implementation of Frame Relay.

companies like Cisco Systems, Sun Microsystems, Netscape (now part of AOL), and Microsoft, IP has surpassed the critical mass stage. There are some compelling reasons for this. First, IP is the foundation of the Internet and has contributed to its popularity. IP and the applications that employ it are extremely popular and pervasive, as seen in the success of PC-based networks in North America and western Europe and the spread of this technology to other areas of the world. Second, the cost of using IP has dropped to the point that it can compete with any proprietary or more sophisticated open protocol (witness the failure of the OSI suite of protocols, which offer much greater flexibility and security). Third, all computer hardware and software supports it, eliminating the compatibility problems that plagued the widespread use of computer networks in past decades. That IP by itself is not the final answer is affirmed by the long lifetime of ATM and Frame Relay. Other hybrid options include IP Switching using ATM as the underlying protocol, IP Tunneling to create a virtual private network (VPN) carrying a different protocol such as Novell Netware's IPX/SPX over the Internet, and VoIP, which permits conventional phone conversations to be made through the Internet.

This was a brief review of requirements related to data communication networks that may be served through the ground segment. The field is evolving daily, and we can only attempt to foresee basic trends and general approaches. Taking this approach, we now move into a discussion of the specific architectures that are in use at the time of this writing. We can anticipate that new arrangements, including the use of onboard processing repeaters, will add functionality to data communication ground segments, which should increase their penetration into global telecommunication markets.

4.2.3 Hub-Based VSAT Networks

The strength of the VSAT hub network, illustrated in Figure 4.14, lies in its ability to support low-cost VSATs with small antennas and low-power SSPA transmitters. The data communications services are centralized in topology to connect to a host computer located at a company or organization headquarters. VSATs are placed at remote locations that need data, voice, and video services from the hub; there is, by design, little or no communication between and among remotes. Rather, the vast majority of traffic is between VSAT and hub, reflecting that the central location has the data (or requires the data) that the remote computers provide (or display/store).

Interface requirements for VSATs are satisfied by the functionality of the indoor unit and internal port cards used to interface on the user side.

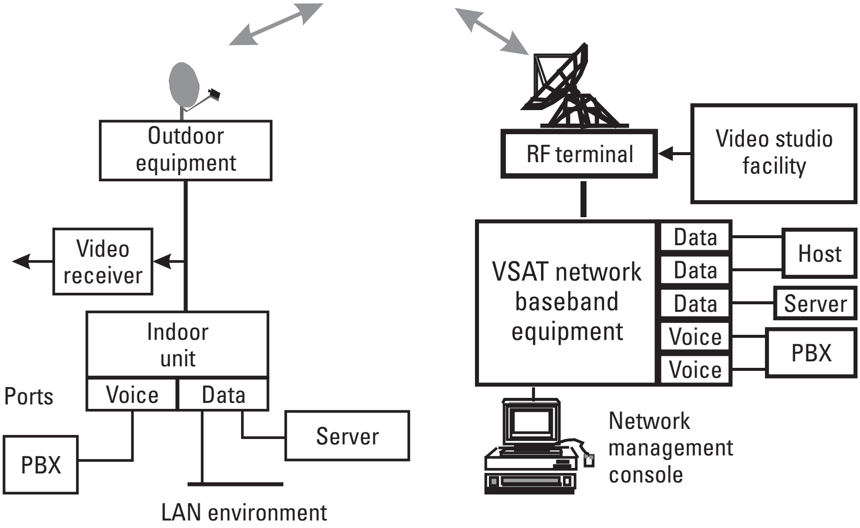


Figure 4.14 Data/voice/video network using a central hub earth station and remote VSAT terminal.

Customer equipment like PCs, a PBX, fax machine, or an IP router within a LAN or WAN environment are typical attachments. Consequently, the VSAT port is designed to accommodate a range of potential connections and to provide the needed spoofing of standard protocols such as TCP/IP and X.25. Within the indoor unit are the multiple access controller and custom modem for the space link. These components form the basic link and physical layers and are proprietary to the manufacturer of the system. Operators must fully understand what protocols need to be supported and that the VSAT network meets this need. Similarly, the hub must have appropriate protocol support for the host computer, which often is connected to it with a terrestrial data line of some distance. Operation of the end-to-end system is centralized to satisfy the demands of user applications and the layers of the protocol stack.

The use of the space segment at Ku band is indicated by the link budget for a typical outroute and inroute in Table 4.7. Hub characteristics include a 6-m transmit/receive parabolic antenna, providing an outroute EIRP of 68.9 dBW and a downlink G/T of 34.7 dB/K. The VSAT transmits its inroute with an EIRP of 48.3 dBW, using a 5-watt SSPA connected to a 1.2-m antenna (the actual power at the antenna transmit port is 3.5 watts, after waveguide and coupling losses). The receive G/T of the VSAT is

Table 4.7

Link Characteristics and Link Budget for a Typical Ku-band Hub-based VSAT Network

	Outroute	**Inroute**	**Units**
Uplink			
Uplink antenna	6.0	1.2	meters
Antenna gain	56.9	42.9	dBi
Transmit power	12.0	5.4	dBW
Uplink EIRP	68.9	48.3	dBW
Path losses	207.0	207.0	dB
Spacecraft G/T	2.0	2.0	dB/K
C/T_{up}	–136.1	–156.7	dBW/K
K	–228.6	–228.6	dB/W/Hz
Bandwidth	500.0	100.0	kHz
IndB	57.0	50.0	dB(Hz)
C/N_{up}	35.5	21.9	dB
Downlink			
Downlink EIRP	25.0	4.4	dBW
Path losses	205.7	205.7	dB
Downlink antenna	1.2	6.0	meters
Antenna gain	41.5	55.5	dBi
System noise temp	120.0	120.0	K
Earth station G/T	20.7	34.7	dB/K
C/T_{dn}	–160.0	–166.6	dBW/K
Bandwidth	57.0	50.0	dB(Hz)
C/N_{dn}	11.7	12.0	dB
Combined Link			
C/N_{th}	11.6	11.6	dB
C/IM	18.0	18.0	dB
C/I	15.0	15.0	dB
C/N_{tot}	9.3	9.3	dB
Threshold	7.0	7.0	dB
Margin	2.3	2.3	dB

20.7 dB/K, which is obtained from a low-cost 80K LNB. Losses and gains are calculated in accordance with the procedure discussed in Chapter 2. The link budget indicates that thermal noise, transponder intermodulation, and RF interference from adjacent satellites combine to produce an overall C/N of 9.3 dB in the outroute and inroute directions, which are both about 2 dB above threshold. This should provide adequate margin for rain fade in most temperate climates.

The outroute and inroute in this example of a typical hub VSAT network are unbalanced because their respective data rates are 512 kbps and 128 kbps, respectively. This means that there would have to be multiple inroutes to create a balance of information transfer between the hub and each VSAT. The exact quantity of inroutes will depend on the amount of inbound data throughput to the hub (which would be more or less than the outbound data, depending on the application). Techniques for evaluating and specifying these components can be found in [18]. The link budget does not include a reserve of hub power for uplink power control, which may be used to compensate for uplink fade (which affects reception at all of the VSATs). Satellite capacity for this network would be reserved in a transponder that contains multiple carriers, e.g., single channel per carrier (SCPC) or, alternatively, a partial transponder.

4.2.4 Mesh VSAT Networks

Mesh topology achieves direct point-to-point connection between pairs of locations with minimum propagation delay and satellite capacity utilization (in terms of the number of links if not power) with the mesh type of topology. For many ground segments, the mesh topology is also attractive because of its flexibility for connecting arbitrary pairs of earth stations. The same arrangement may allow any single earth station to broadcast data to all others. The mesh has been implemented in past years using any of the available MA techniques. As discussed in Chapter 1, mesh networks involving large earth stations have been in use since the 1970s for demand-assigned telephone services. Telephone service imposes severe requirements on the design of the network in terms of call setup times, blocking of calls, bit rate, and service management.

There are fundamentally three classes of connectivity requirements for mesh networks using VSATs. These are:

- *Permanently assigned* (dedicated bandwidth) links for CBR data transfer. This is analogous to a dedicated data line or digital private line using some multiple of 64 kbps, including T1 and E1 services. Higher data rates from DS3 and OC3 are feasible but would put the links out of the class of VSATs under discussion. An example of the latter type of application is to allow direct connection from a remote location to an Internet service provider (ISP), in situations where terrestrial lines of appropriate bandwidth are either not available or too expensive.

- *Demand-assigned* (circuit switched) links between locations for CBR data transfer. Again, the link must not require greater EIRP and G/T from the VSATs than is feasible. A telephony type of VSAT can also offer connection-oriented data services.
- *Dynamic allocation* of bandwidth (packet switched), allowing VBR services on a mesh network basis. This is the same type of facility provided by hub VSAT networks, with the exception that transmissions can be received by any earth station. With the introduction of Ka-band payloads and onboard processing, VSATs of this type will provide a wide range of services using a high-speed protocol such as ATM.

The attractiveness of VSAT mesh networks will be a factor in the growth of ground segments and satellite systems in coming years. At the time of this writing, non-GEO mobile satellite systems had just been introduced. These systems are based on extremely low-cost ultrasmall earth stations. While they provide many special benefits of true portability and operation literally anywhere, they are limited as to data transfer. This is where fixed VSATs hold the greatest promise for medium and high data rate services.

References

[1] Freeman, Roger L., *Telecommunications Transmission Handbook*, 4th ed., New York: Wiley, 1998, p. 50.

[2] The Communications Handbook, CRC Press, with Permission. 1996 (their Figure 27 [Fig.4.3]; their Figure 27.4 [Fig. 404]; their Figure 27.5 [Fig. 4.5]).

[3] ADSL Forum, www.adsl.com/pressroom/dsl_flavors.html.

[4] Rathgeb, Erwin P., "Integrated Services Digital Network (ISDN) and Broadband (B-ISDN)," *The Communications Handbook*, ed. Jerry D. Gibson, Boca Raton, FL: IEEE Press/CRC Press, 1997, p. 581.

[5] "Telephone Transmission Quality, Series P Recommendations," *CCITT,* Vol. 5, 9th Plenary Assembly, Melbourne (location of meeting), 14–25 November 1988; International Telecommunication Union, Geneva (home of ITU), 1989.

[6] Elbert, Bruce R., *The Satellite Communication Applications Handbook*, Norwood, MA: Artech House, 1997, p. 400.

[7] Schwartz, Mischa, *Telecommunications Networks: Protocols, Modeling and Analysis*, Reading, MA: Addison-Wesley, 1987.

[8] Erlang, A. K., "The Theory of Probabilities and Telephone Conversations," *Nyt Tidsskrift for Matematik B*, Vol. 20, 1909.

[9] Boucher, James R., *Voice Teletraffic Systems Engineering*, Dedham, MA: Artech House, 1988, p. 46.

[10] Bartholomew, Martin F., *Successful Business Strategies Using Telecommunications Services*, Norwood, MA: Artech House, 1997.

[11] Aidarous, Salah, and Thomas Plevyak, *Telecommunications Network Management: Technologies and Implementations,* Piscataway, NJ: IEEE Press, 1998, p. 165.

[12] Macario, R. C. V., *Modern Personal Radio Systems*, London: Institution of Electrical Engineers, 1996, p. 137.

[13] Freeman, *Telecommunications Transmission Handbook*, p. 233.

[14] Elbert, Bruce R., *International Telecommunication Management*, Norwood: Artech House, 1990.

[15] Noll, A. Michael, *Introduction to Telephones and Telephone Systems*, 3d ed., Norwood, MA: Artech House, 1999, p. 203.

[16] Elbert, Bruce R., and Bobby Martyna, *Client/Server Computing: Architecture, Applications and Distributed Systems Management,* Norwood, MA: Artech House, 1994.

[17] Wilder, Floyd, *A Guide to the TCP/IP Protocol Suite*, Norwood, MA: Artech House, 1998.

[18] Elbert, *The Satellite Communication Applications Handbook*, Norwood, MA: Artech House, 1997. p. 301.