

1

Introduction

This chapter aims to introduce the basic concepts of satellite networking including applications and services, circuit and packet switching, broadband networks, network protocols and reference models, characteristics of satellite networks, internetworking between satellite and terrestrial networks and convergence of network technologies and protocols. It also highlights the recent development of broadband satellite systems and standards. When you have completed this chapter, you should be able to:

- Understand the concepts of satellite networks and internetworking with terrestrial networks.
- Know the different satellite services, network services and quality of service (QoS).
- Appreciate the differences between satellite networking and terrestrial networking issues.
- Describe the functions of network user terminals and satellite terminals and gateway.
- Know the basic principles of protocols and the ISO reference model.
- Know the basic ATM reference model.
- Know the basic Internet TCP/IP protocol suite.
- Understand the basic concepts of multiplexing and multiple accessing.
- Understand the basic switching concepts including circuit switching, virtual circuit switching and routing.
- Understand the evolution process and convergence of network technologies and protocols.
- Be aware of the recent development of broadband satellite systems and standards.

1.1 Applications and Services of Satellite Networks

Satellites are manmade stars in the sky, and are often mistaken for real stars. To many people, they are full of mystery. Scientists and engineers love to give life to them by calling them birds – like birds, they fly but so high that other creatures can only dream of. They watch the earth from the sky at its vantage point, help us to find our way around the world, carry our telephone calls, emails and web pages, and relay TV programmes across the sky. Actually,

the altitudes of satellites are far beyond the reach of any real bird. When satellites are used for networking, their high altitude enables them to play a unique role in the global network infrastructure (GNI).

Satellite networking is an expanding field, which has developed significantly since the birth of the first telecommunication satellite, from traditional telephony and TV broadcast services to modern broadband and Internet networks, video on demand and digital satellite broadcasts. Many of the technological advances in networking areas are centred on satellite networking. With increasing bandwidth and mobility demands in the horizon, satellite is a logical option to provide greater bandwidth with global coverage beyond the reach of terrestrial networks, and shows great promise for the future. With the development of networking technologies, satellite networks are becoming more and more integrated into the GNI. Integration of mobile ad hoc networks provides new systems for emergency and rescues services with great mobility. Therefore, internetworking with terrestrial networks and protocols is an important part of satellite networking.

The ultimate goal of satellite networking is to provide services and applications. User terminals provide services and applications directly to users. The network provides transportation services to carry information between users' terminals across the network nodes, switches and routers for a certain distance. Figure 1.1 illustrates a typical satellite network configuration consisting of terrestrial networks, satellites with an inter-satellite link (ISL), fixed earth stations, transportable earth stations, portable and handheld terminals, and user terminals connecting to satellite links directly or through terrestrial networks.

1.1.1 Roles of Satellite Networks

In terrestrial networks, many links and nodes are needed to reach long distances and cover wide areas. They are organised to achieve cost effective deployment, maintenance and operation of the networks. The nature of satellites makes them fundamentally different from terrestrial

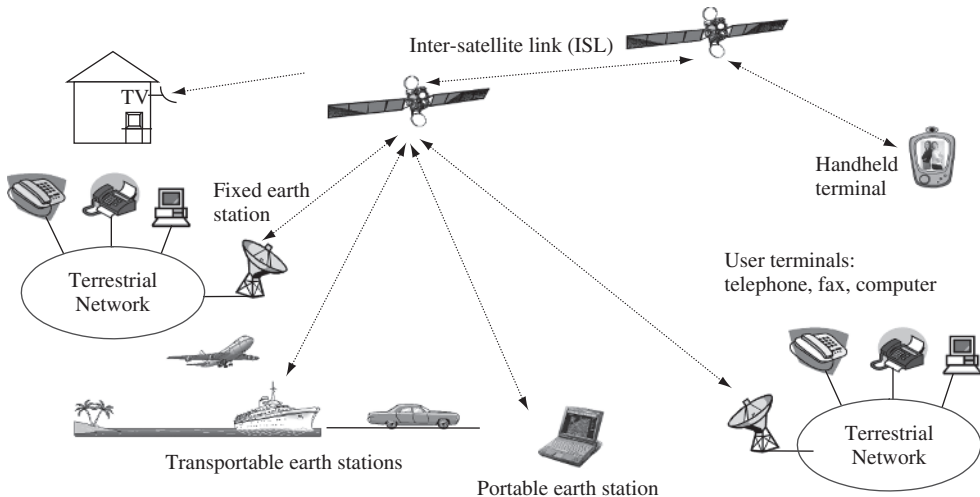


Figure 1.1 Typical applications and services of satellite networking

networks in terms of distances, shared bandwidth resources, transmission technologies, design, development and operation, and costs and applications.

Functionally, satellite networks can provide direct connections among user terminals, connections for terminals to access terrestrial networks, and connections between terrestrial networks as backbones. The user terminals provide services and applications to people, which are often independent from satellite networks, that is, the same terminal can be used to access satellite networks as well as terrestrial networks. The satellite terminals, also called earth stations, and are the earth segment of the satellite networks, providing access points to the satellite networks for user terminals via the user earth station (UES) and for terrestrial networks via the gateway earth station (GES). The satellite is the core of satellite networks and also the centre of the networks in terms of both functions and physical connections. Sometimes there are also connections between satellites, particularly for LEO and MEO satellite networks. Figure 1.2 illustrates the relationship between user terminal, terrestrial network and satellite network.

Typically, satellite networks consist of satellites interconnecting a few large GES and many small UES. The small GES are used for direct access by user terminals and the large UES for connecting terrestrial networks. The satellite UES and GES define the boundary of the satellite network. Like other types of networks, users access satellite networks through the boundary. For mobile and transportable terminals, the functions of user terminal and satellite UES can be integrated into a single unit, but for transportable terminals their antennas are distinguishably visible. There are two parts: in-door unit (IDU) and out-door unit (UDU). The IDU contains transmitter and receiver; and ODU contains low noise block (LNB) converter and block up converter (BUC).

The most important roles of satellite networks are to provide access by user terminals and to internetwork with terrestrial networks so that the applications and services provided by terrestrial networks such as telephony, television, broadband access and Internet connections can be extended to places where cable and terrestrial radio cannot economically be installed and maintained. In addition, satellite networks can also bring these services and applications to ships, aircraft, vehicles, space and places beyond the reach of terrestrial networks. Satellites also play important roles in military, meteorology, global positioning systems (GPS), observation of earth and environments, private data and communication services, and future development

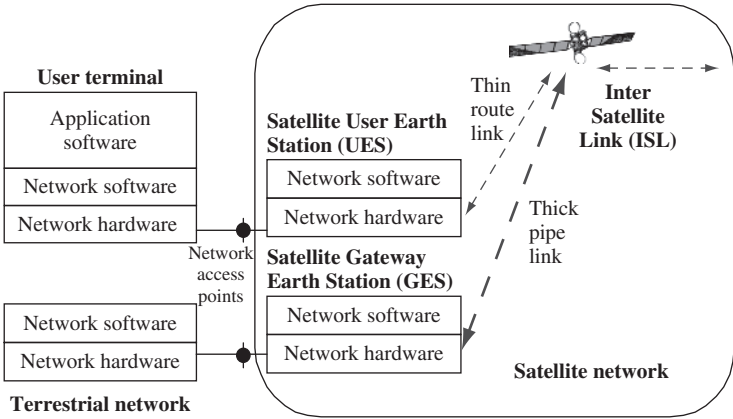


Figure 1.2 Functional relationships of user terminal, terrestrial network and satellite network

of new services and applications for immediate global coverage such as broadband network, and new generations of mobile networks and digital broadcast services worldwide. Emergency and rescue services when disasters strike such as flood, earthquake and volcano.

1.1.2 Network Software and Hardware

In terms of implementation, the user terminal consists of network hardware and software and application software. The network software and hardware provide functions and mechanisms to send information in correct formats using correct protocols at an appropriate network access point. They also receive information from the access point in the same manner.

Network hardware provides signal transmission making efficient and cost-effective use of bandwidth resources and transmission technologies. Naturally, a radio link is used to ease mobility of the user terminals associated with access links; and high-capacity optical fibre is used for backbone connections. Due to propagation environment, radio links are used in satellite system between the earth and satellite segments. There is also research for inter satellite link using optical communications.

With the advance of digital signal processing (DSP), traditional hardware implementations are being replaced more and more by software to increase the flexibility of reconfiguration, hence reducing costs. Therefore the proportion of implementation becomes more and more in software and less and less in hardware. Many hardware implementations are first implemented and emulated in software, though hardware is the foundation of any system implementation.

For example, traditional telephone networks are mainly in hardware; and modern telephone networks, computer, smart-phone, and data networks and the Internet are mainly in software.

1.1.3 Satellite Network Interfaces

Typically, satellite networks have two types of external interfaces: one is between the satellite UES and user terminals; and the other is between the satellite GES and terrestrial networks. Internally, there are three types of interfaces: between the UES and satellite communication payload system; between the GES and satellite communication payload system; and the inter-satellite link (ISL) between satellites. All use radio links, except that the ISL may also use optical links.

Like physical cables, radio bandwidth is one of the most important and scarce resources for information delivery over satellite networks. Unlike cables, bandwidth cannot be manufactured by adding more cables, it can only be shared and its use optimised. The other important resource is transmission power. In particular, power is limited for user terminals requiring mobility or for those installed in remote places that rely on battery supply of power, and also for communication systems on board satellites that rely on battery and solar energy. The bandwidth and transmission power together with the transmission conditions and environment determine the capacity of the satellite networks, as defined by the Shannon Theory.

Satellite networking shares many basic concepts with general networking. In terms of topology, it can be configured into star or mesh topologies. In terms of transmission technology, it can be set up for point-to-point, point-to-multipoint and multipoint-to-multipoint connections. In terms of interface, we can easily map the satellite network into general network terms such as user network interface (UNI) and network nodes interface (NNI).

When two networks need to be connected together, a network-to-network interface is needed, which is the interface of a network node in one network with a network node in another network. They have similar functions as NNI. Therefore, NNI may also be used to denote a network-to-network interface.

1.1.4 Network Services

The UES and GES provide network services. In traditional networks, such services are classified into two categories: teleservices and bearer services. The teleservices are high-level services that can be used by users directly such as telephone, fax service, video, data, email and web services. Quality of service (QoS) at this level is user centric, that is, the QoS indicates users' perceived quality, such as mean objective score (MOS). The bearer services are lower level services provided by the networks to support the teleservices. QoS at this level is network centric, that is, transmission delay, delay jitter, transmission errors and transmission speed. Sometimes, quality of experience (QoE) is assessed by the performance measurements and metrics as well as custom survey.

There are methods to map between these two levels of services. The network needs to allocate resources to meet the QoS requirement and to optimise the network performance. Network QoS and user QoS have contradicting objectives adjustable by traffic loads or network resources, that is, we can increase QoS by reducing traffic load on the network or by increasing network resources; however, this may decrease the network utilisation for network operators. Network operators can also increase network utilisation by increasing traffic load, but this may affect user QoS. It is the art of traffic engineering to optimise network utilisation with a given network load under the condition of meeting user QoS requirements. There is trade-off among the following parameters: traffic load, network resources, QoS, performance and utilization.

1.1.5 Applications

Applications are combinations of one or more network services. For example, tele-education and telemedicine applications are based on combinations of voice, video and data services. Combinations of voice, video and data are also called multimedia services. Some applications can be used with the network services to create new applications.

Services are basic components provided by the networks. Applications are built from these basic components. Often the terms application and service are used interchangeably in the literature. Sometimes it is useful to distinguish them.

1.2 ITU-R Definitions of Satellite Services

Satellite applications are based on the basic satellite services. Due to the nature of radio communications, the satellite services are limited by the available radio frequency bands. Various satellite services have been defined, including fixed satellite service (FSS), mobile satellite service (MSS) and broadcasting satellite service (BSS) by the International Telecommunication Union (ITU) – Radiocommunication Standardisation Sector (ITU-R) for the purpose of bandwidth allocation, planning and management.

1.2.1 Fixed Satellite Service (FSS)

The FSS is defined as a radio communication service between a given position on the earth's surface when one or more satellites are used. These stations at the earth surface are called earth stations of FSS. Stations located on board satellites, mainly consisting of the satellite transponders and associated antennas, are called space stations of the FSS. Of course, new-generation satellites have onboard sophisticated communication systems including onboard switching or routing. Communications between earth stations are through one satellite or more satellites interconnected through ISL. It is also possible to have two satellites interconnected through a common earth station without an ISL. FSS also includes feeder links such as the link between a fixed earth station and satellite for broadcasting satellite service (BSS) or mobile satellite service (MSS). The FSS supports all types of telecommunication and data network services such as telephony, fax, data, video, Internet broadcasting TV and radio.

1.2.2 Mobile Satellite Service (MSS)

The MSS is defined as a radio communication service between mobile earth stations and one or more satellites. This includes maritime, aeronautical and land MSS. Due to mobility requirements, mobile earth terminals are often small, and some are even handheld terminals.

1.2.3 Broadcasting Satellite Service (BSS)

The BSS is a radio communication service in which signals transmitted or retransmitted by satellites are intended for direct reception by the general public using a TV receiving only antenna (TVRO). The satellites implemented for the BSS are often called direct broadcast satellites (DBS). The direct receptions include individual direct to home (DTH) and community antenna television (CATV). The new generation of BSS may also have a return link via satellite.

1.2.4 Other Satellite Services

Some other satellite services are designed for specific applications such as military, radio determination, navigation, meteorology, earth surveys and space exploration. A set of space stations and earth stations working together to provide radio communication is called a satellite system. For convenience, sometimes the satellite system or a part of it is called a satellite network. We will see in the context of network protocols that the satellite system may not need to support all the layers of functions of the protocol stack (physical layer, link layer or network layer).

1.3 ITU-T Definitions of Network Services

During the process of developing broadband communication network standards, the ITU Telecommunication Standardisation Sector (ITU-T) has defined telecommunication services provided to users by networks. These services are provided by user terminals. There are two main classes of services: interactive and distribution services, which are further divided into subclasses.

1.3.1 Interactive Services

Interactive services offer one user the possibility to interact with another user in real-time conversation and messages or to interact with information servers using computers, laptops or mobile phones. It can be seen that different services may have different QoS and bandwidth requirements from the network to support these services. The subclasses of the interactive services are defined as the following:

- *Conversational services*: conversational services in general provide the means for bidirectional communication with real-time (no store-and-forward) end-to-end information transfer from user to user or between user and host (e.g. for data processing). The flow of the user information may be bidirectional symmetric, bidirectional asymmetric and in some specific cases (e.g. such as video surveillance), the flow of information may be unidirectional. The information is generated by the sending user or users, and is dedicated to one or more of the communication partners at the receiving site. Examples of broadband conversational services are telephony, videotelephony and videoconference.
- *Messaging services*: messaging services offer user-to-user communication between individual users via storage units with store-and-forward, mailbox and/or message handling (e.g. information editing, processing and conversion) functions. Examples of broadband messaging services are message-handling services and mail services for moving pictures (videos), high-resolution images and audio information. These also include short message and multimedia message in mobile phones.
- *Retrieval services*: the user of retrieval services can retrieve information stored in information centres provided for public use. This information will be sent to the user by demand only. The information can be retrieved on an individual basis. Moreover, the time at which an information sequence starts is under the control of the user. Examples are broadband retrieval services for videos, high-resolution images, audio information and archival information. These also include the Internet web services.

1.3.2 Distribution Services

This is modelled on traditional broadcast services and video on demand (VoD) to distribute information to a large number of users. The requirement of bandwidth and QoS are quite different from interactive services. The distribution services are further divided into the following subclasses:

- *Distribution services without user individual presentation control*: these services include broadcast services. They provide a continuous flow of information, which is distributed from a central source to an unlimited number of authorised receivers connected to the network. The user can access this flow of information without the ability to determine at which instant the distribution of a stream of information will be started. The user cannot control the start and order of the presentation of the broadcasted information. Depending on the point of time of the user's access, the information will not be presented from the beginning. Examples are broadcast services for television and radio programmes.
- *Distribution services with user individual presentation control*: services of this class also distribute information from a central source to a large number of users. However, the

information is provided as a sequence of information entities (e.g. frames) with cyclical repetition. These can also be buffered in the local devices or remote servers. So, the user has the ability of individual access to the cyclical distributed information and can control the start and order of presentation or from the storage of local device or remote servers. Due to the cyclical repetition and access to the storage, the information entities selected by the user will always be presented from the beginning or any point selected by the user. One example of such a service is video on demand.

1.4 Internet Services and Applications

Like computers, in recent years the Internet has been developed significantly and the use of it has been extended from research institutes, universities and large organisations into ordinary family homes and small businesses.

The Internet was originally designed to interconnect different types of networks including LANs, MANs and WANs. These networks connect different types of computers together to share resources such as memory, processor power, graphic devices and printers. They can also be used to exchange data and for users to access data stored in any of the computers across the Internet.

Today the Internet is not only capable of supporting data, but also image, voice and video on which different network services and applications can be built such as IP telephony, video-conferencing, tele-education and telemedicine.

Like computers, our laptops and mobile phones are also connected to the Internet, even sensor devices, consumer devices at home are also connected to the Internet, called the Internet of Things (IoT), including alarm, security camera, central heating, cooker, wash machines, gas and electricity meters and so on.

The requirements of new services and applications clearly changed the original objectives of the Internet. Therefore, the Internet is evolving towards a new generation to support not only the traditional computer network services but also real-time user services including telephony. Eventually, this will lead to a convergence of the Internet and telecommunication networks towards the future global network infrastructures of which satellite will play an important part.

1.4.1 World Wide Web (WWW)

The WWW enables a wide range of Internet services and applications including e-commerce, e-business and e-government. It also enables virtual meetings with a new style of work, communication, leisure and lives. The WWW is an application built on top of the Internet, but is not the Internet itself. It can be seen that the basic principle of the Internet has not changed much in the last 40 years, but applications and potential of the Internet have changed significantly, particularly the user terminals, user software, services and applications, and human-computer interface (HCI) from keyboard and mouse to touch screen and voice recognition.

The WWW is a distributed, hypermedia-based Internet information system including browsers for users to request information, servers to provide information and the Internet to transport users' requests from users to servers and information from servers to users. There are also powerful search engines to find information on users behalf, in the same time they also collect any information users putting into the Internet.

The hypertext transfer protocol (HTTP) was created in 1990, at the European Organisation for Nuclear Research (originally called the European Council for Nuclear Research; in French: *Conseil Européen pour la Recherche Nucléaire*; CERN), the European Particle Physics Laboratory in Geneva, Switzerland, as a means for sharing scientific data internationally, instantly and inexpensively. With hypertext a word or phrase can contain a link to other text. To achieve this, the hypertext mark up language (HTML), a subset of general mark up language (GML), is used to enable a link within a web page to point to other pages or files in any server connected to the network. This non-linear, non-hierarchical method of accessing information was a breakthrough in information sharing. It quickly became the major source of traffic on the Internet. There are a wide variety of types of information (text, graphics, sounds, movies, business transactions and social media etc.). It is possible to use the web to access information from almost every server connected to the Internet in world. The basic elements for access to the WWW are:

- HTTP: the protocol used for the WWW to transport web pages.
- URL (uniform resource locator): defines a format to address the unique location of the web page identified by the internet protocol (IP) address of a computer, port number within the computer system and location of the page in the file system.
- HTML: the programming 'tags' added to text documents that turn them into hypertext documents.

In the original WWW, the URL identified a static file. Now it can be a dynamic web page created by the servers according to information provided by users; and it can also be an active web page, which is a piece of program code to be downloaded and run on the user's browser computer when clicked.

1.4.2 File Transfer Protocol (FTP)

Long before the WWW, FTP is an application layer protocol providing a service for transferring files between a local computer and a remote computer. FTP is a specific method used to connect to another Internet site to receive and send files. FTP was developed in the early days of the Internet to copy files from computer to computer using a command line. With the advent of WWW browser software, we no longer need to know FTP commands to copy to and from other computers, as web browsers have integrated the commands into their browser functions.

1.4.3 Telnet

This is one of the earliest Internet services providing text-based access to a remote computer. We can use telnet in a local computer to login to a remote computer over the Internet. Normally, an account is needed in the remote host so that the user can enter the system. After a connection is set up between the local computer and remote computer, it allows users to access the remote computer as if it were a local computer.

Such a feature is called location transparency, that is, the user cannot tell the difference between the responses from the local machine or remote machine. It is called time transparency if the response is so fast that user cannot tell the difference between local machine and remote machine by response time. Transparency is an important feature in distributed information

systems. The concepts are also applicable to the cloud computing today. Users can access the cloud resources such as storage spaces and computing powers without needing to know where these resources are.

1.4.4 Electronic Mail (email)

The email is like our postal system but much quicker and cheaper, transmitting only information without papers or other materials, that is, you can order a pizza through the Internet but cannot receive any delivery from it. The early email allowed only text messages to be sent from one user to another via the Internet. Email can also be sent automatically to a list of addresses. Email has grown over the past 20 years, from a technical tool used by research scientists, to a business tool as common as faxes and letters. It is replacing the faxes and letters. Every day, millions and millions of emails are sent through intranet systems and the Internet. We can also use mailing lists to send an email to groups of people. When an email is sent to a mailing list, the email system distributes the email to the listed group of users. It is also possible to send very large files, audio and video clips by attachments.

The success of email systems also causes problems for the Internet, for example, viruses and junk mail are spread through email, threatening the Internet and the many computers linked to it.

1.4.5 Multicast and Content Distribution

Multicast is a generalised case of broadcast and unicast. It allows distribution of information to multiple receivers via the Internet or intranets. Example applications are content distributions including news services, information on stocks, sports, business, entertainment, technology, weather and more. It also allows real-time video and voice broadcast over Internet. This is an extension to the original design of the Internet.

1.4.6 Voice over Internet Protocol (VoIP)

VoIP is one of the important services with significant development over the Internet. This type of service is real time and is more suitable for traditional telecommunication networks. It is different in many ways from the original Internet service. It has quite different traffic characteristics, QoS requirements and needs of bandwidth and network resources.

Digitised streams of voices are segmented into voice ‘frames’. These frames are encapsulated into a voice packet using a real-time transport protocol (RTP) that allows additional information for real-time service including time stamps to be included. The real-time transport control protocol (RTCP) is designed to carry control and signalling information used for VoIP services.

The RTP packets are put into the user datagram protocol (UDP), which is carried through the Internet by IP packets. The QoS of VoIP depends on network conditions in terms of congestion, transmission errors, jitter and delay. It also depends on the quality and available bandwidth of the network such as the bit error rate and transmission speed.

Though the RTP and RTCP were originally designed to support telephony and voice services, they are not limited to these, as they can also support real-time multimedia services including

video services over IP. By making use of the time-stamp information generated at source by the sender, the receiver is able to synchronise different media streams to reproduce the real-time information.

1.4.7 Domain Name System (DNS)

The DNS is an example of application layer services. It is not normally used by users directly, but is a service used by the other Internet applications. It is an Internet service that translates domain names into IP addresses by computers. Because domain names are alphabetical, they are easier to remember by people. The Internet, however, is really based on IP addresses. Every time you use a domain name, therefore, a DNS service must translate the name into the corresponding IP address. For example, the domain name `www.surrey.ac.uk` will translate to IP address: `131.227.132.17`. The IP address can also be used directly.

The DNS is, in fact, a distributed hierarchical system in the Internet. If one DNS server does not know how to translate a particular domain name, it asks another one at a higher level, and so on, until the correct IP address is returned.

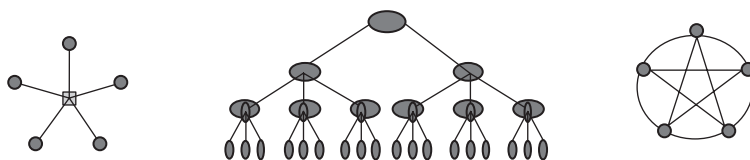
The DNS is organised as a hierarchical distributed database that contains mapping of domain names to various types of information including IP addresses, service type, time to live (TTL) and so on. Therefore, the DNS can also be used to discover other information stored in the database such as mail exchange.

1.5 Circuit-switching Network

The concept of circuit-switching networks comes from the early analogue telephony networks. The network can be of different topologies including star, hierarchical and mesh at different levels to achieve coverage and scalability. Figure 1.3 shows typical topologies of networks.

An example of telephone networks is shown in Figure 1.4. At local exchange (LEX) level, many telephones connect to the exchange forming a star topology (as a complete mesh topology is not scalable). Each trunk exchange (TEX) connects several local exchanges to form the first level of the hierarchy. Depending on the scale of the network, there may be several levels in the hierarchy. At the top level, the number of exchanges is small; therefore a mesh topology is used by adding redundancy to make efficient use of network circuits as well as reliability.

All the telephones have a dedicated link to the local exchange. A circuit is set up when requested by a user dialling the telephone number, which signals the network for a connection.



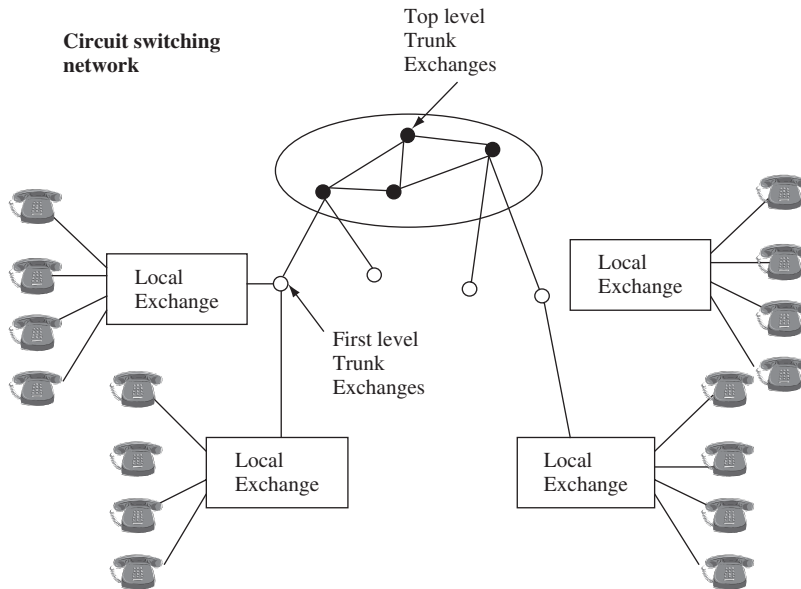


Figure 1.4 Circuit switching networks

1.5.1 Connection Set Up

To set up a connection, a set of circuits has to be connected, joining two telephone sets together. If two telephones are connected to the same LEX, the LEX can set up a circuit directly. Otherwise, additional steps are taken at a higher level TEX to set up a circuit across the switching network to connect to the remote LEX then to the destination telephone.

Each TEX follows routing and signalling procedures. Each telephone is given a unique number or address to identify which LEX it is connected to. The network knows which TEX the LEX is connected to. The off-hook signal when a user picks up the phone and dialled telephone number provide signalling information for the network to find an optimum route to set up a group of circuits to connect the two telephones identified by the calling telephone number and called telephone number.

If the connection is successful, communication can take place, and the connection is closed down after communication has ended that either the caller or callee put down the telephone. If the connection fails or is blocked due to lack of circuits in the network, one has to try again.

At this point, you may imagine that due to the wide coverage of satellite systems, it is possible to have satellites acting as a LEX to connect the telephones directly, or to act as a link to connect LEX to TEX, or connect TEX together. The roles of the satellite in the network have a significant impact on the complexity and cost of the satellite systems, as the different links require different transmission capacities. Satellites can be used for direct connection without strict hierarchy for the scalability needed in terrestrial networks.

The end to end circuit connection is maintained during the communications of the telephony service. Obviously, this is not very efficient for data and Internet services, as transmission times for data is much shorter comparing with the circuit set-up time, hence, it is very expensive to

set up and maintain any connection for the short data service. This is also the reason; all the data networks were designed using connectionless approach, unless there is a good reason to benefit from connection oriented approach for efficiency and reliability.

1.5.2 Signalling

Early generation of switches could only deal with very simple signalling. Signalling information was kept to the minimum and the signal used the same channel as the voice channel.

Modern switches are capable of dealing with a large amount of channels, hence the signalling. The switches themselves have the same processing power as computers, are very flexible and are capable of dealing with data signals. This leads to separation of signal and user traffic, and to the development of common channel signalling (CCS). In CCS schemes, signals are carried by the dedicated signalling channel over a data network, separated from the voice traffic or service channels.

Combination of the flexible computerised switch and CCS enables a better control and management of the telephone network and facilitates new services such as call forwarding, call back and call waiting as well as security.

Signalling between network devices can be very fast, but responses from people are still the same. The processing power of devices can be improved significantly but not people's ability to react. People used to cause stress to network technologies, but now they are often stressed by technologies today.

1.5.3 Transmission Multiplexing Hierarchy based on FDM

Frequency division multiplexing (FDM) is a technique to share bandwidth between different connections in the frequency domain. All transmission systems are design to transmit signals within a bandwidth limit measured in hertz (Hz). The system may allocate a fraction of the bandwidth-called channel to a connection to support a network service such as telephony rather than allocate a physical cable to the connection. This effectively increases the capacity.

When the bandwidth is divided into channels, each channel can support a connection. Therefore, connections from many physical links can be multiplexed into a single physical link with many channels. Similarly, multiplexed connections in one physical connection can be demultiplexed into many physical connections. Figure 1.5 illustrates the concept of multiplexing in the frequency domain.

The given channel can be used to transmit digital as well as analogue signals. However, analogue transmission is more convenient to process in the frequency domain. A traditional telephone channel transmits audio frequency at a bandwidth of 3.1 kHz (from 0.3 to 3.4 kHz). It is transmitted in the form of a single-sideband (SSB) signal with suppressed carriers at 4 kHz spacing. Through multiplexing, 12 or 16 single channels can form a group. Five groups can form a super-group, super-group to master-group or hyper-group, and to super-group and master-group. Figure 1.6 shows the analogue transmission hierarchy.

1.5.4 Transmission Multiplexing Hierarchy based on TDM

Digital signals can be processed conveniently in the time domain. Time division multiplexing (TDM) is a technique to share bandwidth resources in the time domain. A period of time called

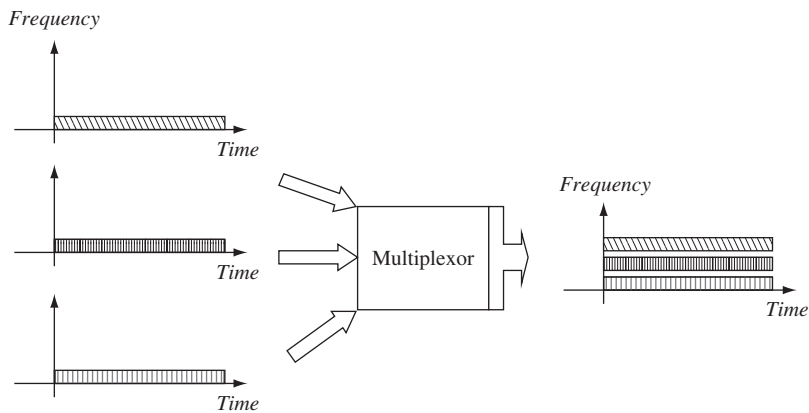


Figure 1.5 Concept of multiplexing in the frequency domain

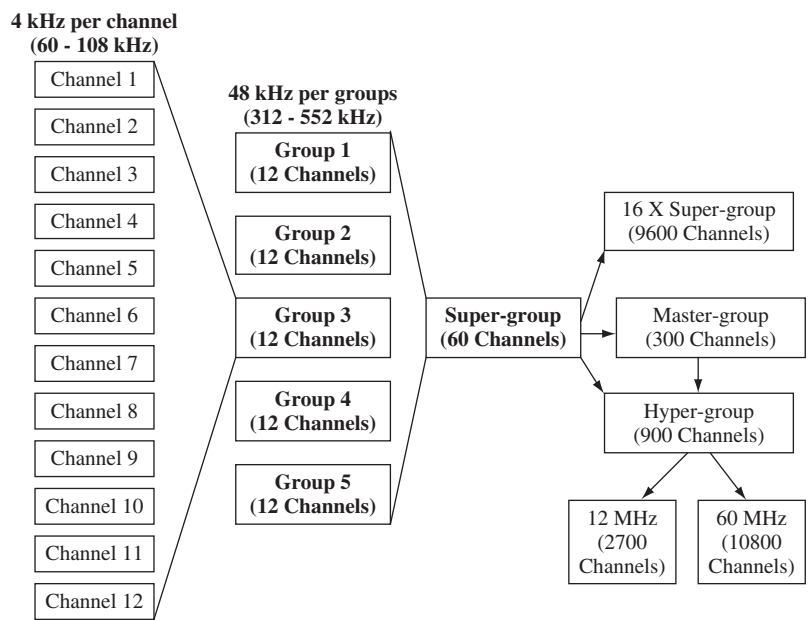


Figure 1.6 Analogue transmission multiplexing hierarchy

a frame can be divided into time slots. Typically, each slot contains a byte or 8 bits digitized information with a duration of 125 μ s. Each time slot can be allocated to a connection. The frame can support the same number of connections as the number of slots. For example, the basic digital connection for telephony is 64 kbit/s. Each byte will take 125 ms to transmit. If the transmission speed is very fast, each byte can be transmitted in a fraction of the 125 ms, and then a time frame of 125 ms can be divided into more time slots to support one connection

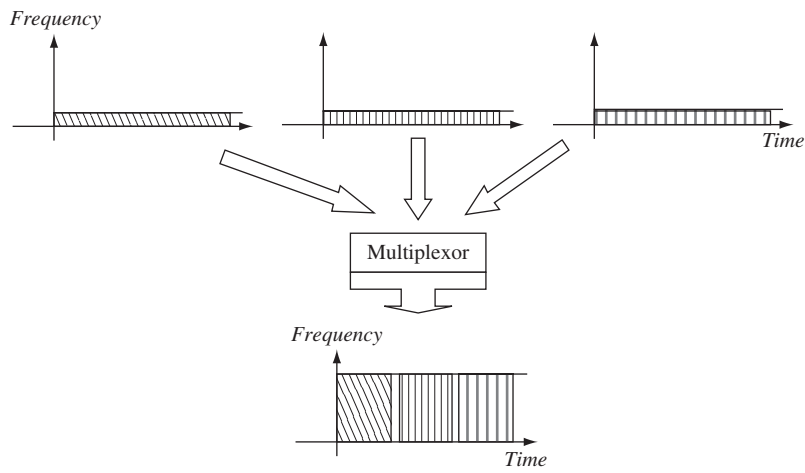


Figure 1.7 Concept of multiplexing in the time domain

for each slot. Several slow bit streams can be multiplexed into one high-speed bit stream. Figure 1.7 illustrates the concept of multiplexing in the time domain.

The digital streams in the trunk and access links are organised into the standard digital signal (DS) hierarchy in North America: DS1, DS2, DS3, DS4 and higher levels starting from 1.544 Mbit/s. In Europe, they are organised into E1, E2, E3, E4 and higher levels starting from 2.048 Mbit/s. The two hierarchies can only internetwork at certain levels, however, the basic rate is the same 64 kbit/s needed to accommodate one telephone circuit. Additional bits or bytes are added to the multiplexed bit stream for signalling and synchronisation purposes, which are also different between North America and European systems. Figure 1.8 shows the transmission multiplexing hierarchies.

1.5.5 Space Switching and Time Switching

In telephony networks and broadcasting networks, the usage of each channel normally is in the order of minutes or hours. The requirements for bandwidth resources are also well defined. For example, channels for telephony services and broadcast services are all well defined.

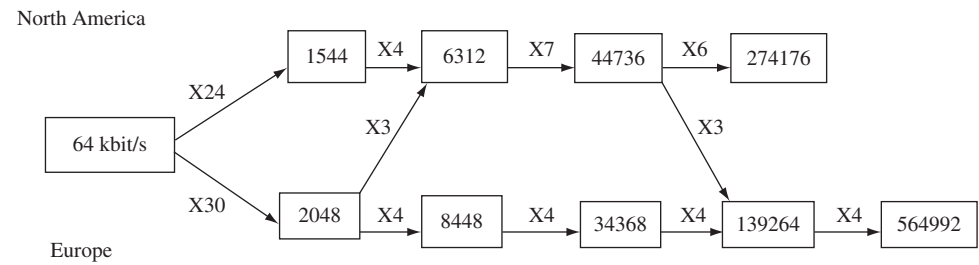


Figure 1.8 Digital transmission hierarchies

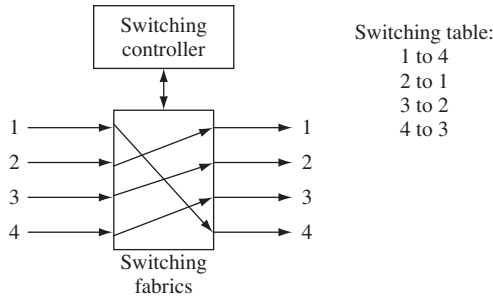


Figure 1.9 Space switching concept

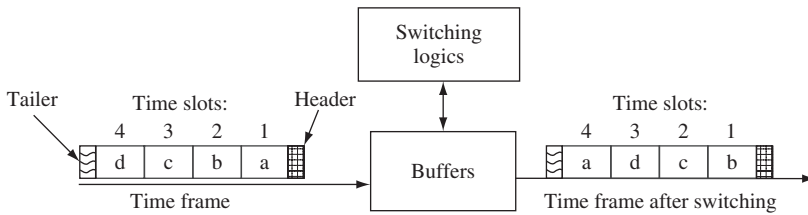


Figure 1.10 Time switching concept

If a switch cannot buffer any information, space in terms of bandwidth or time slots has to be reserved to allow information to flow and switched across the switch as shown in Figure 1.9. This means that the switch can only perform space switching.

If a switch can buffer a frame of time slots, the output of slot contents in the frame can be switched sending out the contents in a different order, as shown in Figure 1.10. This means that the switch can perform time switching. Switch designs can use either/or a combination of space switching and time switching, such as space-time-space or time-space-time combinations.

1.5.6 Coding Gain of Forward Error Correction (FEC)

In satellite networking, the transmission from satellite to the earth station is normally power limited. To make it worse, there may be propagation loss and increased noise power. Therefore, it is important to introduce an error correction coding, that is, to add additional information to the data so that some errors can be corrected by the receiver. This is called forward error correction (FEC), because the additional information and processing take place before any error occurs.

Depending on modulation schemes, bit error probability (BEP) is expressed as a function of E_b/N_0 which is related to E_c/N_0 by the expression:

$$E_b/N_0 = E_c/N_0 - 10 \log \rho \quad (1.1)$$

where E_b is the energy per bit without coding, E_c is the energy per bit with coding, N_0 is the noise spectral density (W/Hz) and $\rho = n/(n + r)$ is the code rate (where r is the number of bits

added for n information bits). It can be seen that we can use less power to improve the BEP at the cost of additional bits (hence bandwidth). The value $(10 \log \rho)$ is called the coding gain. There is also a trade-off between power and bandwidth for a given BEP.

Using $C = E_c R_c$, we calculate:

$$E_c/N_0 = (C/R_c)/N_0 = (C/N_0)/R_c \quad (1.2)$$

where C is carrier power and R_c is the channel bit rate. It can be seen that the same BEP can be achieved using less transmission power, at the cost of added redundant information for error correction.

1.6 Packet-switching Networks

The packet switching concept was developed for computer networks, because streams of bits or bytes do not make much sense to computers. The computer needs to know the start and end of the data transmission.

In a data network, it is important to be able to identify where transmission of data starts and where transmission ends. The data, together with identifiers of the start and end of the data, is called a frame. In addition, addresses, frame checks and other information are added so that the sending computer can tell the receiving computer what to do based on a protocol or rule sent when the frame is received. If the frame is exchanged on a link between two computers, it is defined by the link layer protocol. The frame is a special packet on links. Therefore, the frame is related to link layer functions.

Information can also be added to the frame to create a packet so that the computer can make use of it to route the packet from the source to the destination across the network when more network links and nodes are needed. Therefore, the packet is related to network layer functions.

The initial packet network was designed for transmission of messages or data. The start and end of the data, correctness of transmission and mechanisms to detect and recover errors are all important. If the communication channel is perfect, a complete message can be handled efficiently as a whole, however, in the real world, this assumption cannot be met easily. Therefore, it is practical to break down the large message into smaller segments using packets for transmission. If there is any error in the message, only the error packet needs to be dealt with rather than the whole message retransmitted.

With packets, we do not need to divide bandwidth resources into narrow channels or small time slots to meet service requirement. We can use the complete bandwidth resources to transmit packets at high speed. If we need more bandwidth, we can simply use more or larger packets to send our data. If we use less bandwidth, we use fewer and smaller packets. Packets provide flexibility for bandwidth resource allocations, particularly when we do not know the requirement of bandwidth resources from some new multimedia services.

The meaning of broadband has been defined by the ITU-T as a system or transmission capable of dealing with data rates higher than the primary rates, which are 1.544 Mbit/s in North America and 2.048 Mbit/s in Europe. Today, data rates have reached 100 Mbit/s even 10 Gbit/s in terrestrial networks with optical fibre. In satellite and wireless networks, researchers and engineers are striving to achieve the same capacity as the terrestrial networks to be comparable in both access and transit networks.

There are two approaches for implementation of the packet-switching network. One is used in traditional telephony networks and the other is used in the computer and data networks.

1.6.1 Connection-oriented Approach

In a packet-switching network, each physical connection has a much wider bandwidth, which is capable of supporting high-speed data transmissions. To divide this bandwidth for more connections, the concept of a virtual channel is used. The packet header carries an identification number to identify different logical connections within the same physical connection.

On receiving the packet, the packet switch can forward the packet to the next switch using another virtual channel until the packet reaches its destination. For switching, the network needs to be set up before the packet is transmitted. That is, a switching table needs to be set up in the switch to connect the incoming virtual channels to the outgoing virtual channels. If connection requirements are known, the network can reserve resources for the virtual connections in terms of packets and their payload.

This approach is called the virtual channel approach. Like telephony networks, the virtual channel based approach is connection oriented, that is, a connection needs to be set up before communication taking place. All packets follow the same connection from source to destination. The connection is called virtual connection.

In circuit switching, physical paths are set up to switch from input channels to output channels. In virtual channel switching, channels are identified by logic numbers; hence changing the logic number identifier virtually switches the packets to a different logical channel. Virtual channel switching is also called virtual circuit switching. Figure 1.11 illustrates the concept of virtual channel switching.

Some virtual channels can be grouped logically to form a virtual path, just like the group in the circuit networks. The network node is called a packet switch, and functions like traditional circuit switching, but it gives flexibility of allocating different amounts of resources to each virtual connection. Therefore it is a useful concept for a broadband network, and is used in

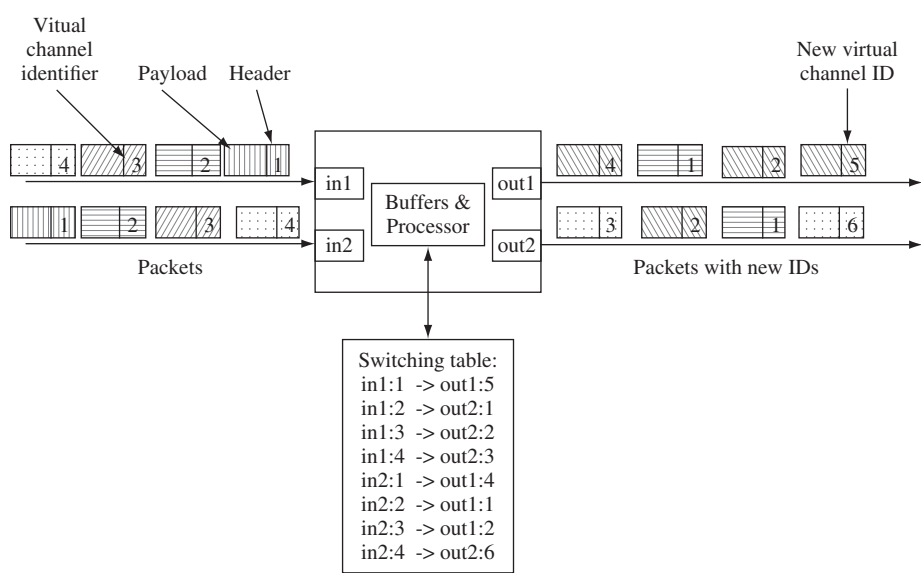


Figure 1.11 Virtual channel switching concept

the asynchronous transfer mode (ATM) network. The virtual connection identifiers are only significant to each switch for identifying logical channels.

This kind of network is quite similar to our telephony and railway networks. Resources in term of times can be reserved to guarantee QoS during the connection set-up stage. The network blocks the connection request if there are not enough resources to accommodate the additional connection.

1.6.2 Connectionless Approach

In computer and data networks, transmission of information often takes a very short period of time compared to the time setting up the telephone connections. It becomes inefficient to set up a connection for the computer and data networks for each packet transmission.

To overcome the problem with the virtual channel approach, the connectionless approach is used to transmit packets from sources to destinations without pre-setting connections. Such a packet is called the datagram because it consists of source and destination addresses rather than connection identifiers to allow the network node (also called the router) to route the packet from source to destination. Figure 1.12 illustrates the concept of connectionless approach.

Such network called connectionless networks. In a connectionless network, the packet header needs to carry the destination address so that the network can use it to route the packet from source to destination, and also the source address for response by the destination computer. The network packet switch is called a router to distinguish it from the connection-oriented switch or traditional channel-based switch. The router has a routing table containing information about destination and the next node leading to the destination with minimum costs.

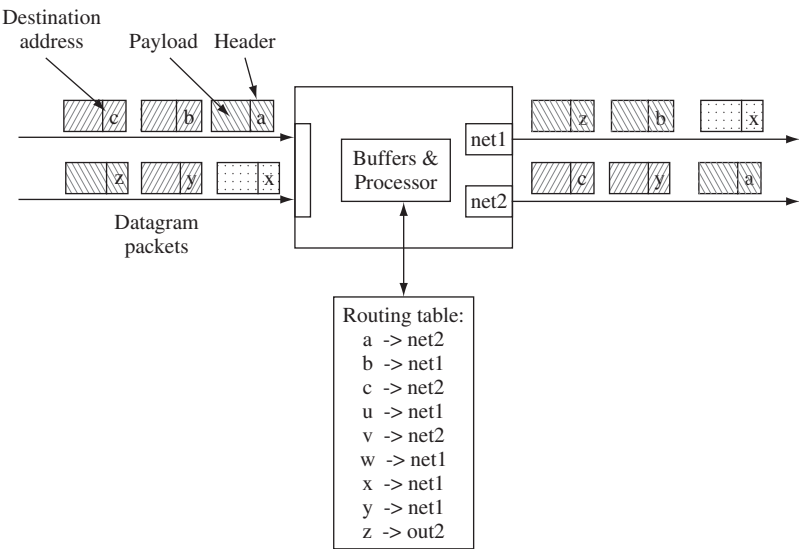


Figure 1.12 Datagram routing concept

The connectionless approach has flexibility for individual packets to change to different routes if there is congestion or failure in the route to destination. This kind of network is quite similar to postal delivery or motorway networks in the United Kingdom. There is no way to make a reservation beforehand; hence, there is no guarantee of QoS. When traffic conditions are good, one car journey can give a good estimate of travel time. Otherwise, it may take much more time to reach the destination and sometimes it can be too late to be useful. However, there is flexibility to change its route after starting the journey to avoid any congestion or closure in the route. The Internet is an example of this kind of network, hence the information highway is a high-speed delivery network; hence, a good description of the information infrastructure widely used today.

1.6.3 Relationship between Circuit-switching and Packet-switching

Circuit switching relates more closely to transmission technologies than packet switching. It provides physical transmission of signals carrying information in the networks. The signals can be analogue and digital. For analogue signals it provides bandwidth resources in term of Hz, kHz or MHz, treated in a frequency domain such as frequency division multiplexing (FDM). For digital signals it provides bandwidth resources in term of bit/s, kbit/s or Mbit/s, even reaching Gbit/s and Tbit/s today, treated in a time domain such as time division multiplexing (TDM). It is also possible to take into account both time and frequency domains such as code division multiple access (CDMA). At circuit level, switches deal with streams of bits and bytes of digital signals to flow along the circuits or analogue signals with defined bandwidth. There is no structure in the signal.

Packets provide a level of abstraction above the bit or byte level, by providing structure to bit streams. Each packet consists of a header and payload. The header carries information to be used by the network for processing, signalling, switching and controlling purposes. The payload carries user information to be sent, received and processed by user terminals.

On top of a circuit it is possible to transmit packets. With packets it is possible to emulate the circuit by continuous streams of packets. This is how the voice over Internet protocol (VoIP) and broadcasting streaming services are implemented by continuous flow of IP packets. These allow internetworking between circuit networks and packet networks. The emulated circuit is called a virtual circuit. It can be seen that virtual circuit, frame and packet are different levels of abstract from physical transmissions to network layer functions.

1.6.4 Considerations of Packet Network Designs

Packet is a layer of functions introduced to the networks. It separates the user services and applications from transmission technologies. A packet provides flexibility for carrying different types of data such as voice, video and web services without involving transmission technologies and media. The network deals with packets only rather than different services and applications. The packets can be carried by any types of networks including satellite.

Introducing packets into networks brings tremendous benefit for developing new services and applications and for exploring new network technologies, and also brings a great challenge to network designers.

What size should the packet be? There should be a trade-off between requirements from applications and services and the capabilities of transmission technologies. If is too small, it

may not be capable of meeting the requirements of the services and applications, but if it is too big it may not be fully utilised and may also cause problems in transmission. Large packets are more likely to get bit errors than small ones, as transmission channels are never perfect in real life. For large packets it takes a long time to transmit and process and they also need large memory space to buffer them. Real-time services may not be able to tolerate long delays, hence there is a preference for small packets.

1.6.5 Packet Header and Payload

How many bits should be used for the packet header and how many for payload? With a large header, it is possible to carry more control and signal information. It also allows more bits to be used for addresses for end systems, but it can be very inefficient if services need only a very small payload. There are also special cases for large headers, for example, a large header may be needed for secure transmission of credit card transactions.

1.6.6 Complexity and Heterogeneous Networks

The complexity is due to a large range of services and applications and different transmission technologies. Many different networks have been developed to support a wide range of services and applications and to better utilise bandwidth resources based on packet-switching technologies. Systems may not work together if they are developed with different specifications of packets. Therefore such issues have to be dealt with in a much wider community in order for systems to interwork globally. This is often achieved by developing common international standards. The Internet Protocol (IP) is the best example to internetworking between different network technologies to support all types of Internet services and applications.

1.6.7 Performance of Packet Transmissions

At bit or byte level, transmission errors are overcome by increasing transmission power and/or bandwidth using better channel coding and modulation together with the error detection and error correction techniques. In real systems, it is impossible to eliminate bit errors completely. The errors at bit level will propagate to packet levels. Retransmission mechanisms are used to recover the error/lost packets, thus controlling the error at packet levels. Therefore, packet transmission can be made reliable even if bit level transmissions are unreliable. However, this additional error recovery capability is at the cost of additional transmission time and buffer space. It also relies on efficient error detection schemes and acknowledgement packets to confirm a successful transmission. For the retransmission scheme, the efficiency of channel utilisation can be calculated as:

$$\eta = t_i / (t_i + 2t_p + t_r) \quad (1.3)$$

where t_i is the time for transmission of a packet onto the channels, t_p is the time for propagation of the packet along the channel to the receiver, and t_r is the processing time of the acknowledgement packet by the receiver. It can be seen that large packet transmission times or small propagation times and packet processing times are good for packet transmission performance.

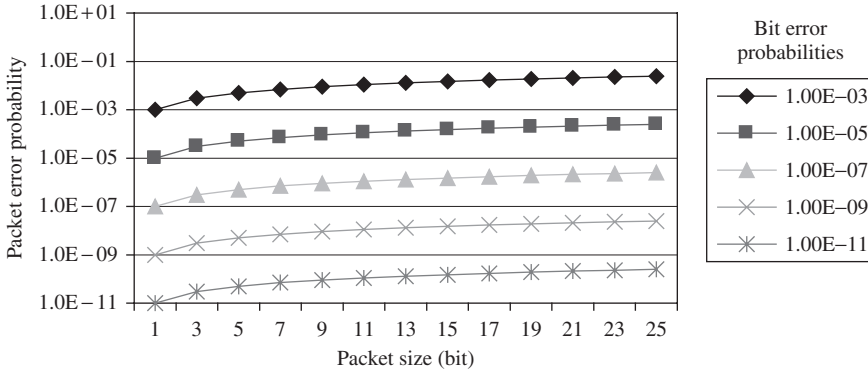


Figure 1.13 Packet error probabilities for given bit error probabilities and packet sizes

1.6.8 Impact of Bit Level Errors on Packet Level

We may quickly realise that the probability of packet error is related to the packet size and quality of bit level transmissions. If P_b is the probability of error at bit level, the probability of packet error P_p of n bits can be calculated as:

$$P_p = 1 - (1 - P_b)^n \quad (1.4)$$

where it is assumed that the error pattern is random at the bit level. If the error pattern is bursty, inter-leaving technique can be applied to randomized the error pattern. Then Equation 1.4 can be used as a good approximation. Figure 1.13 shows the packet error probabilities for given bit error probabilities and packet sizes.

1.7 OSI/ISO Reference Model

Protocols are important for communications between entities. There are many options available to set protocols. For global communications, protocols are important to be internationally acceptable. Obviously, the International Standards Organisation (ISO) has played a very important role in setting and standardising a reference model so that any implementations following the reference model will be able to internetwork and communicate with each other.

Like any international protocol, it is easy to agree in principle how to define the reference model but always difficult to agree about details such as how many layers the model should have, how many bytes a packet should have, how many headers a packet should have to accommodate more functionalities but minimise overheads, whether to provide best-effort or guaranteed services, whether to provide connection-oriented services or connectionless services and so on.

There are endless possible options and trade-offs with many technological selections and political considerations. Without losing generality, here we explain only very briefly the basic concepts and principles, as well as the evolution of the reference models.

1.7.1 Protocol Terminology

A protocol is the rules and conventions used in conversation by agreement between the communicating parties. A reference model provides all the roles so that all parties will be able to communicate with each other if they follow the roles defined in the reference model in their implementation.

To reduce design complexity, the whole functions of systems and protocols are divided into layers, and each layer is designed to offer certain services to higher layers, shielding those layers from the details of how the services are actually implemented.

Each layer has an interface with the primitive operations, which can be used to access the offered services. Network protocol architecture is a set of layers and protocols.

A protocol stack is a list of protocols (one protocol per layer). An entity is the active element in each layer, such as user terminals, switches and routers. Peer entities are the entities in the same layer capable of communication with the same protocols.

Basic protocol functions include segmentation and reassembly, encapsulation, connection control, ordered delivery, flow control, error control and routing and multiplexing.

Protocols are needed to enable communicating parties to understand each other and make sense of received information. International standards are important to achieve a global acceptance. Protocols described in the standards are often in the context of reference models, as many different standards have been developed.

1.7.2 Layering Principle

The layering principle is an important concept for network protocols and reference models. In the 1980s, the ISO derived the seven-layer reference model shown in Figure 1.14, called the open systems interconnection (OSI) reference model, which is based on clear and simple principles.

It is the first complete reference model developed as an international standard. The principles that were applied to arrive at the seven layers can be summarised as:

- A layer defines a level of abstraction which should be a different from any other layer.
- Each layer performs a well-defined function.
- The function of each layer should be chosen to lead to internationally standardised protocols.
- The layer boundaries should be chosen to minimise information flow across the interface.
- The number of layers should be large enough to accommodate all the distinguish functions but not too large to be too complicated.

1.7.3 Functions of the Seven Layers

The following are brief descriptions of the functions of each layer in the OSI/OSI seven-layer reference model.

- Layer 1 – the physical layer (bit stream) specifies mechanical, electrical and procedure interfaces and the physical transmission medium. In satellite networks, radio links are the physical transmission media; modulation and channel coding enable the bit stream to be transmitted in defined signals and allocated frequency bands.

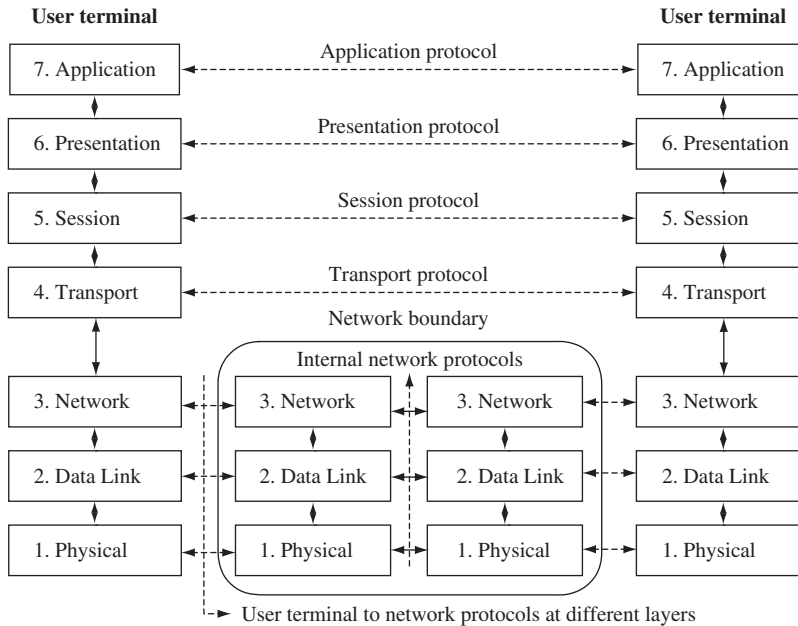


Figure 1.14 OSI/ISO seven-layer reference model

- Layer 2 – the data link layer provides a line between network entities that appears free of undetected transmission errors to the network layer. Broadcasting media have additional issues in data link layer, that is, how to control access to the shared medium. A special sublayer called the medium access control (MAC) scheme, such as Polling, Aloha, FDMA, TDMA, CDMA, DAMA, deals with this problem.
- Layer 3 – the network layer routes packets from source to destination. The functions include network addressing, congestion control, accounting, disassembling and reassembling, coping with heterogeneous network protocols and technologies. In broadcast networks, the routing problem is simple: the routing protocol is often thin or even non-existent.
- Layer 4 – the transport layer provides a reliable data delivery service for high-layer users between the user end devices. The transport layer is the highest layer of the services associated with the provider of communication services. The higher layers are user data services. It has functions of ordered delivery, error control, flow control and congestion control.
- Layer 5 – the session layer provides the means of cooperating presentation entities to organise and synchronise their dialogue and to manage the data exchange in the user end devices.
- Layer 6 – the presentation layers are concerned with data transformation, data formatting and data syntax.
- Layer 7 – the application layer is the highest layer of the ISO architecture. It provides services to application processes, such as email, file transfer and web services.

1.7.4 Fading of the OSI/ISO Reference Model

We have seen the rapid development of many types of new applications, services, networks and transmission media. No one expected such a fast development of the Internet and new

services and applications. New technologies and new service and application developments have changed the conditions of the optimisation points of the layering functions as one of the reasons leading to the fading of the international standards.

There are also many other reasons, including technical, political and economical reasons, or too complicated to be used in a practical world. The reference model is not much used in today's networks. However, the layering principles are still widely used in network protocol design and implementation. It is the classical and true reference model that all modern protocols always try to use as a reference to discuss and describe the functions of their protocols and evaluate their performance by analysis, simulation and experiment.

1.8 The ATM Protocol Reference Model

The asynchronous transfer model (ATM) is based on fast packet switching techniques for the integration of telecommunications and computer networks. Historically, telephone networks and data networks were developed independently. Development of integrated services digital network (ISDN) standards by the ITU-T was the first attempt to integrate telephony and data networks. It was based on 64 kbit/s channels, was called narrowband ISDN (N-ISDN), as the broadband ISDN (B-ISDN) was developed as soon as the N-ISDN finished.

1.8.1 *Narrowband ISDN (N-ISDN)*

N-ISDN provides two 64 kbit/s digital channels, which replace the analogue telephone services, plus a 16 kbit/s data channel for signalling and data services from homes to local exchanges. The ISDN follows the concept of circuit networks very closely, as the envisaged main services at the time, telephony and high-speed data transfer, need no more than 64 kbit/s. The primary rates are 1.5 Mbit/s for North America and 2 Mbit/s for Europe.

1.8.2 *Broadband ISDN (B-ISDN)*

ATM is a further effort by ITU-T to develop a broadband integrated services digital network (B-ISDN) following the development of ISDN, which is called narrowband ISDN (N-ISDN) to distinguish it from B-ISDN.

As soon as standardisation of the N-ISDN was complete, it was realized by the international community that the N-ISDN based on circuit networks could not meet the increasing demand by new services, applications and data networks such as HDTV and video conferences.

The standardisation processes of B-ISDN led to the development of ATM based on the concept of packet switching. It provides flexibility of allocating bandwidth to user services and applications from tens of kbit/s used for telephony services to hundreds of Mbit/s for high-speed data and high definition TV (HDTV).

The ITU-T recommended that the ATM is the target solution for broadband ISDN. It is the first time in its history that international standards were set up before development.

1.8.3 *ATM Technology*

The basic ATM technology is very simple. It is based on a fixed packet size of 53 bytes of which 5 bytes are for the header and 48 for the payload. The ATM packet is called a "cell" as the basic transport unit of the broadband networks, due to its small and fixed size.

It is based on the virtual channel switching approach providing a connection-oriented service and allowing negotiation of bandwidth resources and QoS for different applications. It also provides control and management functions to manage the systems, traffic and services for generating revenue from the network operations.

1.8.4 Reference Model

The reference model covers three plans: user, control and management. All transportation aspects are in the form of ATM, as shown in Figure 1.15, including:

- The physical layer provides physical media-related transmissions such as optical, electrical and microwave.
- The ATM layer defines ATM cells and related ATM functions.
- The ATM adaptation layer adapts high-layer protocols including the services and applications and divides data into small segments so that they can be suitable for transportation in the ATM cells.

1.8.5 Problems: Lack of Available Services and Applications

The ATM has been influenced by the development of optical fibre, which provides very large bandwidths and very low transmission errors. However, there are a large number of network technologies developed to support the IP including wired and wireless links.

Services and applications are considered as parts of functions in user terminals rather than as parts of the network. The networks are designed to be able to meet all the requirements of services and applications. However, the higher layers were never defined and so few services and applications were developed on the ATM network. ATM has tried to inter-network with all different sorts of networks including some legacy networks together with the management and control functions making ATM very complicated and expensive to implement.

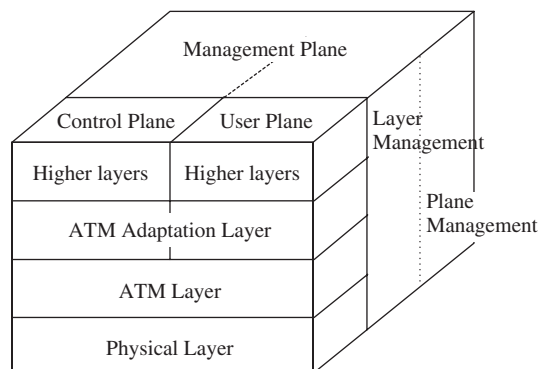


Figure 1.15 B-ISDN ATM reference model

1.9 Internet Protocols Reference Model

Originally, the Internet protocols (IP) were not developed by any international standardisation organisation. They were developed by the United States Department of Defense (DoD) research project funded by the Defense Advanced Research Project Agency (DARPA) to connect a number of different networks designed by different vendors into a network of networks (the ‘Internet’). It was initially successful because it delivered a few basic services that everyone needed (file transfer, electronic mail, telnet for remote logon) across a very large number of different systems.

The main part of the Internet protocol reference model is the suite of transmission control protocol (TCP) and Internet protocol (IP) known as the TCP/IP protocols. Several computers in a small department can use TCP/IP (along with other protocols) on a single LAN or a few interconnected LANs. The Internet protocols allow the construction of very large networks with less central management.

As all other communications protocol, TCP/IP is composed of different layers but is much simpler than the OSI/ISO and ATM. Figure 1.16 shows the Internet reference model.

1.9.1 Network Layer: IP Protocol

The network layer is the Internet protocol (IP) based on the datagram approach, proving only best effort service without any guarantee of quality of service (QoS). IP is responsible for moving packets of data from node to node. IP forwards each packet based on a four-byte destination address (the IP address). The Internet authorities assign ranges of numbers to different organisations. The organisations assign groups of their numbers to departments. On 6 June 1012, it started officially the evolution process from the IPv4 toward next generation IP called IPv6.

1.9.2 Network Technologies

The network technologies, including satellite networks, LANs, ATM and so on, are not part of the protocols. They transport IP packets from one edge of the network to the other edge.

Layers (OSI)

Application	HTTP	SMTP	FTP	Telnet	DNS	RTP/RTCP	etc....
Transport	TCP / UDP						
----- Network -----	IP						
Link + Physical	Satellite Network	Wireless LAN	Ethernet LAN	ATM Network	etc....		

Figure 1.16 The Internet reference model

The source host sends IP packets and the destination host receives the packets. The network nodes route the IP packets to the next routers or gateways until they can route the packets directly to the destination hosts. All these network technologies converge towards Ip, including telephony, computer, broadcasting and mobile communication networks.

1.9.3 Transport Layer: TCP and UDP

The transmission control protocol (TCP) and the user datagram protocol (UDP) are transport layer protocols of the Internet protocol reference model. They provide ports or sockets for services and applications at user terminals to send and receive data for different services and applications across the Internet.

The TCP is responsible for verifying the correct delivery of data between client and server. Data can be lost in the intermediate network. TCP adds support to detect errors or lost data and to trigger retransmission until the data is correctly and completely received. Therefore TCP provides a reliable service though the network underneath may be unreliable, that is, operation of Internet protocols do not require reliable transmission of packets, but reliable transmission can reduce the number of retransmissions and hence increase performance.

UDP provides the best-effort service without trying to recover any error or loss. Therefore, it is also a protocol providing unreliable transport of user data. However, this is very useful for real-time application, as retransmission of any packet may cause more problems than the lost packets.

1.9.4 Application Layer

The application layer protocols are designed as functions of the user terminals or servers. The classical Internet application layer protocols include HTTP for WWW, FTP for file transfer, SMTP for email, telnet for remote login, DNS for domain name service and more, including real-time protocol (RTP), real-time control protocol (RTCP) and real time streaming protocol (RTSP) for real-time services, and others for dynamic and active web services. All these should be independent from the networks.

1.9.5 QoS and Control on Resources

Most functions of the Internet define the high layer protocols. The original Internet protocol version 4 (IPv4) provides only best-effort services; hence it does not support any control functions and cannot provide any quality of services. Extensions have been made during the recent years including QoS, security, mobility, multicast and address space. The problems are addressed in the next generation of the Internet protocol version 6 (IPv6). Research is also carried out for the future Internet design beyond the IPv6.

1.10 Satellite Network

There are two types of transmission technologies: broadcast and point-to-point transmissions. Satellite networks can support both broadcast and point-to-point connections. Satellite networks are most useful where the properties of broadcast and wide coverage are important.

Satellite networking plays an important role in providing global coverage complementing to the terrestrial networks. There are three types of roles that satellites can play in communication networks: access network, transit network and broadcast network.

1.10.1 Access Network

The access network provides access for user terminals or private local networks. Historically in telephony networks, it provided connections from telephone or private branch exchanges (PBX) to the telephony networks. The user terminals link to the satellite earth terminals to access satellite links directly. Today, in addition to the telephony access network, the access networks can also be broadband Internet access. As an example, a laptop can be connected to the Internet for email and web services via a broadband satellite system.

1.10.2 Transit Network

The transit network provides connection between networks or network switches. It often has a large capacity to support a large number of connections for large volume of network traffic. Users do not have direct access to it. Therefore they are often transparent to users, though they may notice some differences due to propagation delay or quality of the link via a satellite network. Examples of satellite as transit networks include interconnect international telephony networks and broadband Internet backbone networks. Bandwidth sharing is often pre-planned using fixed assignment multiple access (FAMA).

1.10.3 Broadcast Network

Satellite supports both telecommunication service and broadcast service. Satellite can provide very efficient broadcasting services including digital audio and video broadcast (DVB-S) and DVB with return channels via satellite (DVB-RCS). New generation of the standards are also developed, called DVB-S2 and DVB-RCS2 as the second generation of the satellite DVB standards.

1.10.4 Space Segment

The main components of a communication satellite system consist of the space segment: satellite together with the related management and control systems, and the ground segment: earth stations to interconnection with the user terminals and terrestrial networks (see Figure 1.17). The design of satellite networks is concerned with service requirements, orbit and coverage and frequency band selection.

The satellite is the core of the satellite network consisting of a communication subsystem and platform. The platform, also called a bus, provides the structure support and power supply of the communication subsystems, and also includes altitude control, orbit control, thermal control, tracking, telemetry and telecommand (TT&T) to maintain normal operations of the satellite system.

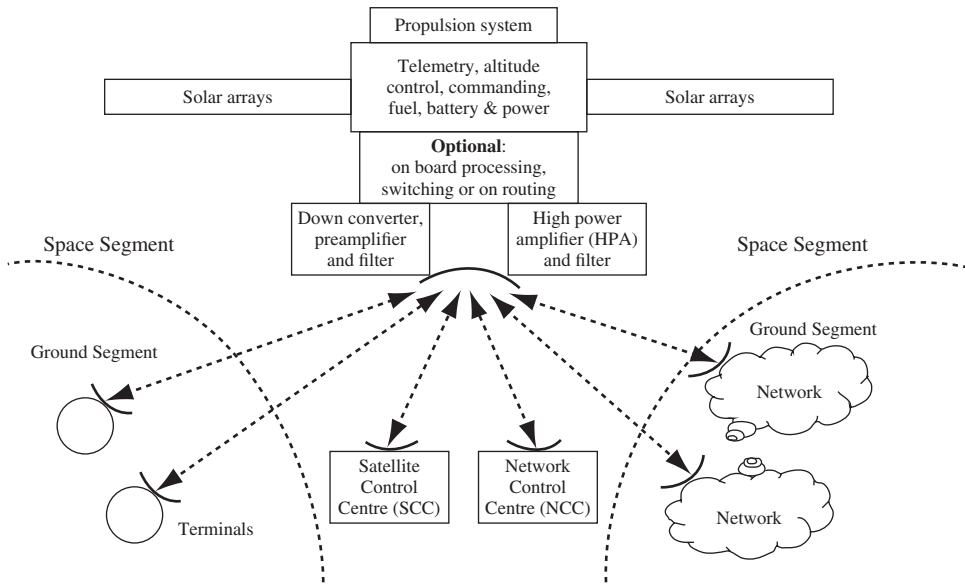


Figure 1.17 Illustration of the space segment and ground segment

The telecommunication subsystems consist of transponders and antenna. The antennas associated with the transponders are specially designed to provide coverage for the satellite network. Modern satellites may also have onboard processing (OBP) and onboard switching (OBS). There are different types of on board functions:

- Transparent transponders provide the function of relaying radio signals. They receive transmissions from the earth station and retransmit them to the earth station after amplification and frequency translation. Satellites with transparent transponders are called transparent satellites.
- OBP transponders provide addition functions including digital signal processing (DSP), regeneration and base band signal processing before retransmitting the signal from satellite to the earth station. Satellites with OBP transponders are called OBP satellites.
- OBS transponders have additional functions than OBP transponders, providing switching functions. Similarly, satellites with OBS transponders are called OBS satellites. With the rapid development of the Internet, experiments are also in progress to fly onboard IP routers.

In addition, the satellite control centre (SCC) and network control centre (NCC) or network management centre (NMC), are parts of the space segment, though they are located at ground level:

- **Satellite control centre (SCC):** this is the on-ground system responsible for the operation of the satellite. It monitors the status of the different satellite subsystems through telemetry links, controls the satellite on its nominal orbit through telecommand links. It communicates with the satellite using dedicated links, which are different from the communication links. It normally consists of typically one earth station and GEO or non-GEO satellite systems,

receiving telemetry from the satellites and sending telecommands to the satellites. Sometimes, a backup centre is built at a different location to improve reliability and availability.

- Network control centre (NCC) or network management centre (NMC): this has different functions from the SCC managing the communication systems on board the satellite. Its main functions are to manage the network traffic and associated resources on board the satellite and on ground to achieve efficient use of the satellite network for communications.

1.10.5 Ground Segment

Ground segment contains user earth stations and gateway earth stations. The earth stations are part of the satellite network. It provides functions of transmitting and receiving traffic signals to and from satellites. It also provides interfaces to terrestrial networks or to user terminals directly. The earth station may consist of the following parts:

- The transmitting and receiving antenna are the most visible parts of the earth station. There are different sizes typically ranging from less than 0.5 to greater than 16 m. The small ones are normally used user terminals and large ones for gateway station.
- Low noise amplifier of the receiver system with a noise temperature ranging from about 30 K to a few hundred K.
- High performance amplifier (HPA) of the transmitter with power from a few watts to a few thousand kilowatts, depending on capacity.
- Modulation, demodulation and frequency translation.
- Signal processing.
- Interfaces to terrestrial networks or user terminals.

1.10.6 Satellite Orbits

Orbits are one of the importance resources for satellite in space, as satellites need to be in a right orbit to provide coverage to the service areas. There are different ways to classify satellite orbits (see Figure 1.18).

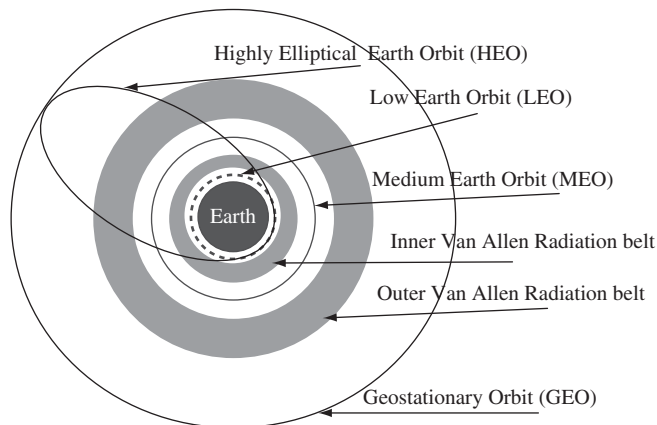


Figure 1.18 Satellite orbits

According to the altitude of satellites, satellite orbits can be classified as the following types:

- Low earth orbit (LEO) has an altitude range of less than 5000 km. Satellites in this type of orbit are called LEO satellites. The period of the satellite is about 2–4 hour (h).
- Media earth orbit (MEO) has an altitude range between 5000 and 20 000 km. Satellites in this type of orbit are called MEO satellites. The period of the satellite is about 4–12 h.
- Highly elliptical earth orbit (HEO) has an altitude range of more than 20 000 km. Satellites in this type of orbit are called HEO satellites. The period of the satellite is more than 12 h.
- Geo-synchronous orbit (GSO) has semi-major axis of 42 164 km (26 199 miles). The period of the satellite is about 24 h (to be more precise, 23 h 56 min 4 s).
- Geostationary orbit (GEO) is a circular geosynchronous orbit in the plane of the Earth's equator with a radius of approximately 42 164 km (26 199 miles) measured from the centre of the Earth. It has an altitude of approximately 35 786 km or 22 236 miles above mean sea level. It maintains the same position relative to the Earth's surface.

Please note that the space surrounding the earth is not as empty as it looks. There are mainly two kinds of space environment constraints to be considered when choosing orbit altitude.

- The Van Allen radiation belts where energetic particles such as protons and electrons are confined by the earth's magnetic field. They can cause damage to the electronic and electrical components of the satellite.
- Space debris belts where spacecraft are abandoned at end of their lifetime. They are becoming of increasing concern to the international community as they can also cause damage to satellite networks particularly satellite constellations and to space missions in the future.

1.10.7 Satellite Transmission Frequency Bands

Frequency bandwidth is another important resource of satellite networking and also a scarce resource. The radio frequency spectrum extends from about 3 kHz to 300 GHz; communications above 60 GHz are generally not practical because of the high power needed and equipment costs. Parts of this bandwidth are used for terrestrial microwave communication links historically, and for terrestrial mobile communications such as GSM and 3/4G networks and wireless LANs today.

In addition, the propagation environment between the satellite and earth station due to rain, snow, gas and other factors and limited satellite power from solar and battery limits further suitable bandwidth for satellite communications. Figure 1.19 shows attenuations of different frequency bands due to rain, fog and gas.

Link capacity is limited by both the bandwidth and transmission power used for transmission. Frequency bandwidths are allocated by the ITU. There are several bands allocated

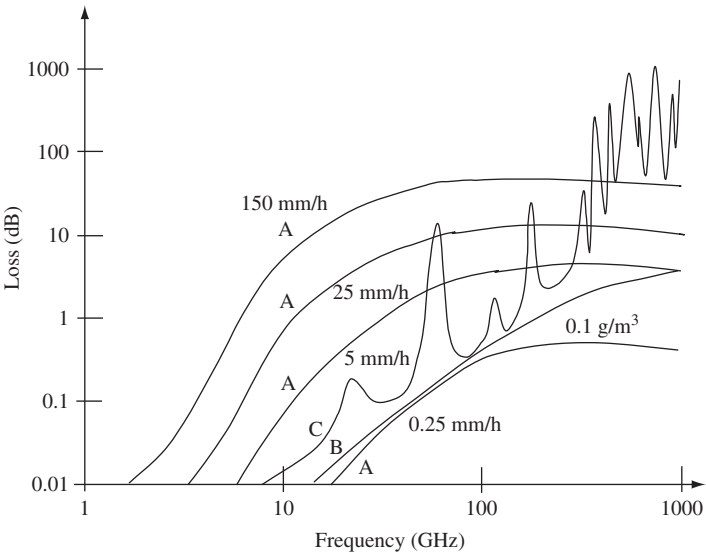


Figure 1.19 Attenuations of different frequency bands due to rain (A), fog (B) and gas (C)

Table 1.1 Typical frequency bands of satellite communications

Denomination	Frequency bands (GHz)
UHF	0.3–1.0
L band	1.0–2.0
S band	2.0–4.0
C band	4.0–8.0
X band	8.0–12.0
Ku band	12.0–18.0
K band	18.0–27.0
Ka band	27.0–40.0

for satellite communications. Table 1.1 shows the different available bandwidths for satellite communications.

Historically, bandwidths around 6 GHz for uplink and 4 GHz for downlink (i.e., 6/4 GHz) have been commonly paired in the C band. Many FSS still use these bands. Military and governmental systems use bands around 8/7 GHz in the X band. There are also some systems that operate around 14/12 GHz in the Ku band. New-generation broadband satellites are using the Ka band to explore wide bandwidth due to saturation of the Ku band and its capacity to support broadband communications. Table 1.2 gives examples of uses of frequency bands.

Table 1.2 Example usages of frequency bands for GEO

Denomination	Uplink (bandwidth)	Downlink (bandwidth)	Typical utilisation in FSS for GEO
6/4 C band	5.850–6.425 (575 MHz)	3.625–4.2 (575 MHz)	International and domestic satellites: Intelsat, USA, Canada, China, France, Japan, Indonesia
8/7 X band	7.925–8.425 (500 MHz)	7.25–7.75 (500 MHz)	Governmental and military satellites
		10.95–11.2	International and domestic satellites in Region 1 and 3
		11.45–11.7	
		12.5–12.75 (1000 MHz)	Intelsat, Eutelsat, France, German, Spain, Russia
13–14/11–12 Ku band	13.75–14.5 (750 MHz)		
		10.95–11.2	International and domestic satellites in Region 2
		11.45–11.7	
		12.5–12.75 (700 MHz)	Intelsat, USA, Canada, Spain
18/12	17.3–18.1 (800 MHz)	BSS bands	Feeder link for BSS
30/20 Ka band	27.5–30.0 (2500 MHz)	17.7–20.2 (2500 MHz)	International and domestic satellites Europe, USA, Japan
40/20 Ka band	42.5–45.5 (3000 MHz)	18.2, 21.2 (3000 MHz)	Governmental and military satellites

1.11 Characteristics of Satellite Networks

Most of the presently employed communication satellites are radio frequency (RF) repeaters, often called “bent pipe” satellites. A processing satellite, as a minimum, regenerates the received digital signal. It may decode and recode a digital bit stream. It also may have some bulk switching capability and for LEO satellite constellation the inter satellite links (ISL).

Radio link (microwave LOS) provides real transmission of the bits and bytes at the physical layer of the layered reference model. There are three basic technical problems in the satellite radio link due to the satellite being located at great distances from the user earth stations.

1.11.1 Propagation Delay

The first problem to deal with is very long distances. For GEO satellites, the time required to traverse these distances – namely, earth station to satellite and then to another earth station – is in the order of 250 ms, depending on the location of the satellite and earth station. The round-trip delay will be of 2×250 or 500 ms. These propagation times are much greater than those encountered in conventional terrestrial systems. One of the major problems is propagation time and resulting echo on telephone circuits. It delays the reply of certain data circuits for block

or packet transmission systems and requires careful selection of telephone signalling systems, or call set-up time may become excessive. Even for packet networks, such a long delay can degrade the QoS of telephony, video conference and web services.

1.11.2 Propagation Loss and Power Limited

The second problem is that there are far greater losses. For LOS microwave we encounter free-space losses possibly as high as 145 dB. In the case of a satellite with a range 35 786 km or 22 236 miles operating on 4.2 GHz, the free-space loss is 196 dB and at 6 GHz, 199 dB. At 14 GHz the loss is about 207 dB. This presents an insurmountable problem from earth to satellite, where comparatively high-power transmitters and very high-gain antennas may be used. From satellite to earth, the link is power-limited for two reasons:

1. In bands shared with terrestrial services, such as the popular 4 GHz band, to ensure non-interference with those services.
2. In the satellite itself, which can derive power only from solar cells. It takes a great number of solar cells to produce the RF power necessary; thus the downlink from satellite to earth is critical, and received signal levels will be much lower than on comparative radio links, as low as -150 dBW.

1.11.3 Orbit Space and Bandwidth Limited for Coverage

The third problem is crowding. The equatorial orbit is filling up with geostationary satellites. Radio-frequency interference from one satellite system to another is increasing. This is particularly true for systems employing smaller antennas at earth stations with their inherent wider beam widths. It all boils down to a frequency congestion of emitters. Currently, there are over 400 satellites in the GEO orbit; and it is still increase at an accelerating speed.

1.11.4 Operational Complexity for LEO

In addition to the GEO satellite, we also see several new low earth orbit satellite systems in operation, which can explore the potential of satellite capabilities. These satellites typically have much lower altitude orbits above the earth. This may reduce the problems of delay and loss, but introduce more complexity in maintaining communication links between earth terminals and satellites due to the fast movement of LEO constellation satellites. One of the example is the Iridium satellite constellation which consists of 66 LEO satellites providing voice and data to satellite phones. The new generation called the Iridium NEXT is expecting to launch in 2015.

1.12 Channel Capacity of Digital Transmissions

In the frequency domain, greater bandwidth can support more communication channels and high transmission capacity. In the time domain, the digital transmission capacity is also directly proportional to the bandwidth.

1.12.1 The Nyquist Formula for Noiseless Channels

For a noiseless channel, the Nyquist formula is used to determine the channel capacity:

$$C = 2B \log_2 M \quad (1.5)$$

where C is the maximum channel capacity for data transfer rate (in bit/s), B is bandwidth (in hertz) and M is the number of levels per signalling element.

1.12.2 The Shannon Theorem for Noise Channels

The Shannon and Hartley capacity theorem is used to determine the maximum bit rate C over a band-limited channel giving a specific signal-to-noise (S/N) ratio. The theorem is:

$$C = B \log_2(1 + S/N) \quad (1.6)$$

where C is the maximum capacity (in bit/s), B is bandwidth of the channel, S is signal power and N is noise power.

As $S = (RE_b)$ and $N = (N_0B)$ the formula can be rewritten in a different form as the following:

$$\begin{aligned} C &= B \log_2[1 + RE_b/(N_0B)] \\ &= B \log_2[1 + (R/B)(E_b/N_0)] \end{aligned} \quad (1.7)$$

where E_b is energy per bit, R is transmission bit rate and $N = N_0B$ where N_0 is noise power spectral density.

1.12.3 Channel Capacity Boundary

Let $R = C$ in Equation 1.7, we get the capacity boundary function between bandwidth efficiency C/B and given E_b/N_0 :

$$C/B = \log_2[1 + (C/B)(E_b/N_0)] \quad (1.8)$$

Then:

$$E_b/N_0 = (2^{C/B} - 1)/(C/B) \quad (1.9)$$

Figure 1.20 shows the relationship of the capacity boundary of the communication channel with E_b/N_0 . If the transmission data rate is within the capacity limit, that is, if $R < C$, we may be able to achieve transmission rate with properly designed modulation and coding mechanisms, and if $R > C$, it is impossible to achieve error free transmission.

1.12.4 The Shannon Power Limit (-1.6 dB)

We can increase the bandwidth to reduce transmission power as a trade-off. If we let the transmission bit rate R achieve the maximum, then from Equation 1.8 we can get the following:

$$(E_b/N_0)^{-1} = \log_2[1 + (C/B)(E_b/N_0)]^{(B/C)/(Eb/N_0)} \quad (1.10)$$

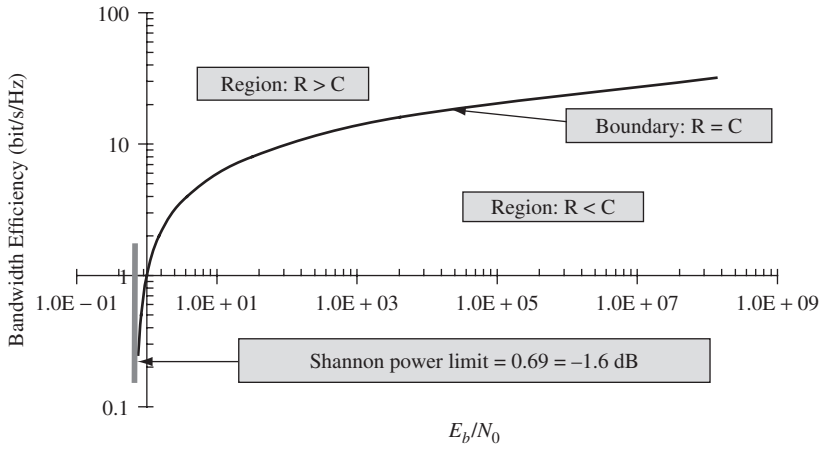


Figure 1.20 Capacity boundary of communication channel

As $(1 + 1/x)^x \rightarrow e$ when $x \rightarrow \infty$, letting $B \rightarrow \infty$ we can get the Shannon power limit:

$$E_b/N_0 = \frac{1}{\log_2(e)} = \log_e 2 \approx 0.69 = -1.6 \text{ dB} \quad (1.11)$$

This tells us, no matter how much bandwidth we have, the transmission power in terms of E_b/N_0 should be larger than the Shannon limit, though there is a trade-off between bandwidth and power.

Given bandwidth B and transmission rate R , we can get from Equation 1.9 that the minimum power requirement for error-free transmission is:

$$E_b/N_0 > (2^{R/B} - 1)/(R/B).$$

1.12.5 Shannon Bandwidth Efficiency for Large E_b/N_0

Similarly we can derive the formula of Shannon bandwidth efficiency from Equation 1.8 for large E_b/N_0 , as the following:

$$\log_2[(C/B)(E_b/N_0)] \leq C/B \leq 1 + \log_2[(C/B)(E_b/N_0)]$$

Hence, $(C/B) \approx \log_2(E_b/N_0)$, when $(E_b/N_0) \rightarrow \infty$

Figure 1.21 shows the convergence between (C/B) and $\log_2(E_b/N_0)$. It also shows that when transmission power is low, increasing the power by a small amount will have a large impact on the bandwidth efficiency; and when transmission power is high reducing bandwidth efficiency by a small amount will have a large saving on transmission power.

Therefore engineers can trade between transmission bandwidth and transmission power, but should not go too far to benefit from such a trade off.

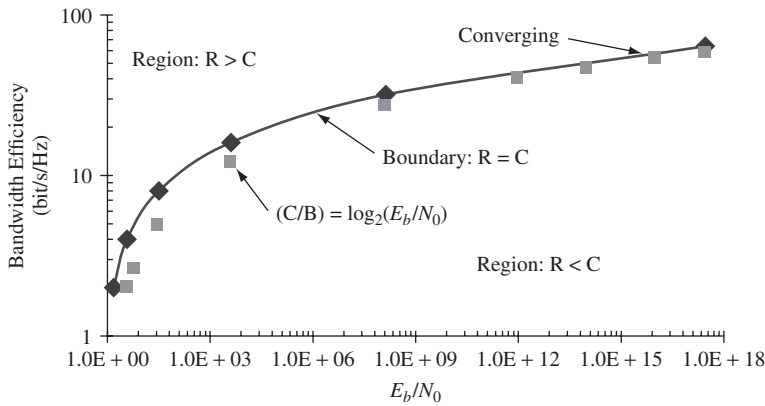


Figure 1.21 Shannon bandwidth efficiency for large E_b/N_0

1.13 Internetworking with Terrestrial Networks

Internetworking techniques are well developed in terrestrial networks. When we have different types of networks we face problems at different layers of the protocol stacks, such as different transmission media, different transmission speeds, different data formats and different protocols. Since networking only involves the lower three layers of the protocols, satellite networking with other types of networks could involve any of the three layers.

1.13.1 Repeaters at the Physical Layer

At the physical layer, internetworking is at the bit level. The internetworking repeater needs to have a function to deal with the digital signal. It is relatively easy to internetwork between the terrestrial network and satellite networks, as the physical layer protocol functions are very simple. The main problem is dealing with data transmission rate mismatch, as terrestrial networks may have much higher data transmission rates.

The main disadvantage of this solution is that it is inflexible due to the nature of implementation at the physical layer. One may have quickly noticed that the communication payload of transparent satellites, relay satellites or bent-pipe satellites deals with bit streams as functions of a repeater.

Most of the commercial devices provide RJ45 port for Ethernet connection between satellite terminal and user terminal. Often USB and telephone interfaces are also provided.

1.13.2 Bridges at the Link Layer

A bridge is a store and forward device and is normally used in the context of LANs, interconnecting one or more LANs at the link layer. In satellite networking, we borrow the term to refer to the internetworking unit between the satellite network and terrestrial networks. As it works at the link layer, it also relies on the physical layer transmission, that is, the bridge deals with the functions of two layers: physical and link layers.

A frame arriving from the satellite network will be checked to decide if the frame should be forwarded to the terrestrial networks according to its routing table and the destination address. If yes, the bridge forwards the frame to the terrestrial networks, otherwise it discards it. Before forwarding, the frame is formatted based on the protocol of the terrestrial networks.

Similar procedures are also carried out when frames flow from the terrestrial networks to the satellite network. The main disadvantage is that the satellite has to deal with a large number of different types of networks and protocol translations. It has more complicated functions than repeaters.

The main advantage is that the satellite network will be able to make use of the link layer functions such as error detection, flow control and frame retransmission. The satellite payload can also implement the bridge functions. Otherwise, the link layer functions have to be carried out on the other side of the satellite networks.

In addition to protocol translations, the satellite terminals or gateways also carry out transmission rate adaptations between the satellite networks and user terminals or terrestrial networks.

1.13.3 Switches at the Physical, Link and Network Layers

Switches can work at any layer of the three layers depending on the nature of the networks. Switching networks can set up end-to-end connections to transport bit streams, frames and even network layer packets.

The main advantage is that switching networks can reserve network resources when setting up connections to meet up QoS requirements.

The disadvantages are that they are not very efficient when dealing with short data transmission and supporting connectionless network protocols such as the Internet, and that it is difficult to deal with heterogeneous networks, but can be more efficient for homogeneous networks.

1.13.4 Routers for Interconnecting Heterogeneous Networks

Router here refers to an Internet router or an IP router. It deals with Internet protocol (IP) packets in addition to physical and link layers. Figure 1.22 shows how routers can be used to internetwork with heterogeneous terrestrial networks. Here it requires that all user terminals use the IP protocol. In the recent development, all network technologies and user terminals as well as network services and applications are evolving towards all IP solutions including mobile, fixed, broadcasting and computer as well as satellite networks.

1.13.5 Protocol Translation, Stacking and Tunnelling

It can be seen that there are three basic techniques for interconnecting heterogeneous networks:

1. Protocol translation: this technique is normally used at the physical layer dependent sublayers of the link layer. Protocol translations are carried out between the different sublayers.
2. Protocol stacking: this technique is normally used for different layers. One layer is stacked on top of the other network.
3. Protocol tunnelling: this technique is similar to protocol staking, but with two of the same type of networks communicating through a tunnelling of other networks.

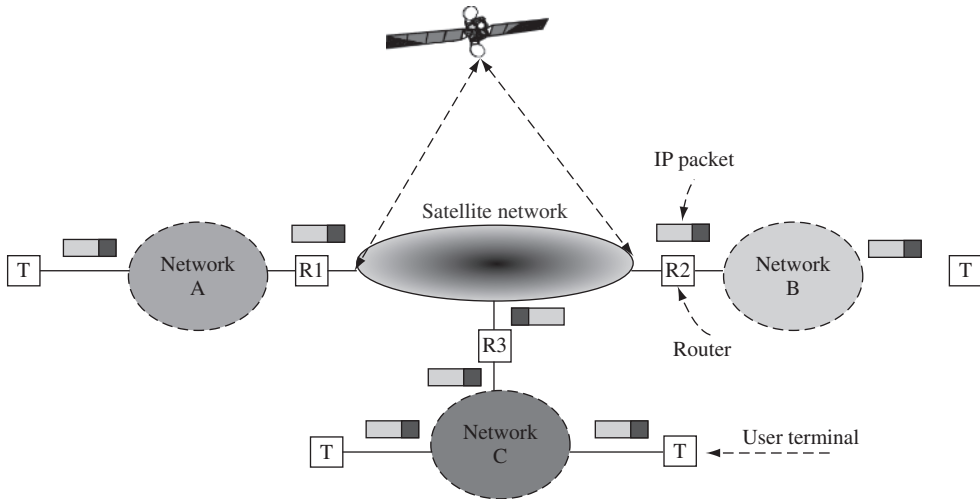


Figure 1.22 Using routers to internetwork with heterogeneous terrestrial networks

1.13.6 Quality of Service (QoS)

The term quality of service (QoS) is extensively used today. It is not only used in analogue and digital transmission in telephony networks but also in broadband networks, wireless networks, multimedia services and even the Internet. Networks and systems are gradually being designed with consideration of the end-to-end performance required by user applications. Most traditional Internet applications such as email and ftp are sensitive to packet loss but can tolerate delays. For multimedia applications (voice and video) this is generally the opposite. They can tolerate some packet loss but are sensitive to delay and variation of the delay.

Therefore, networks should have mechanisms for allocating bandwidth resources to guarantee a specific QoS for real-time applications. QoS can be described as a set of parameters that describes the quality of a specific stream of data provided to the users.

In addition, quality of experience (QoE) becomes important when selling satellite services to users, as they can always compare with the services provided by the terrestrial networks. QoE is often assessed by customer survey.

1.13.7 End-user QoS Class and Requirements

Based on the end-user application requirements, ITU-T recommendation G.1010 defines classification of performance requirements into end-user QoS categories.

Based on the target performance requirements, the various applications can be mapped onto axes of packet loss and one-way delay as shown in Figure 1.23. The size and shape of the boxes provide a general indication of the limit of delay and information loss tolerable for each application class.

It can be seen that there are eight distinct groups, which encompass the range of applications identified. Within these eight groupings there is a primary segregation between applications that can tolerate some information loss and those that cannot tolerate any information loss at all, and four general areas of delay tolerance.

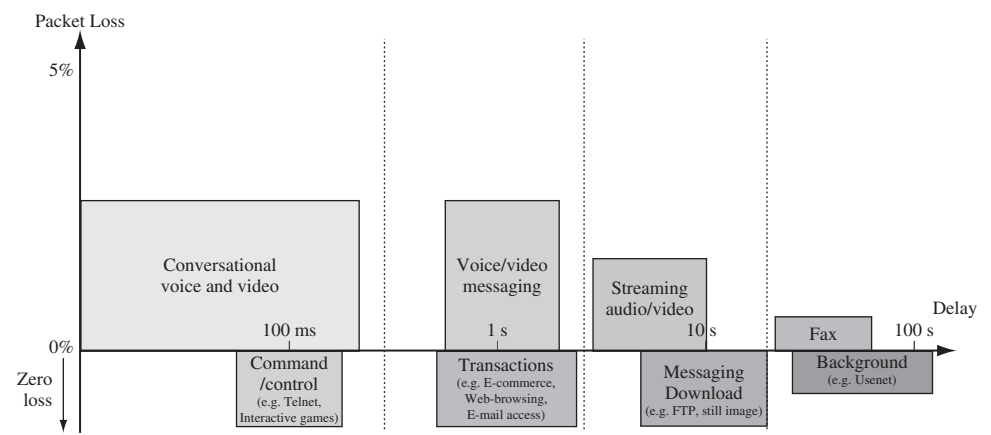


Figure 1.23 Mapping of user-centric QoS requirements into network performance (Source: ITU 2001 [18]. Reproduced with permission of ITU.)

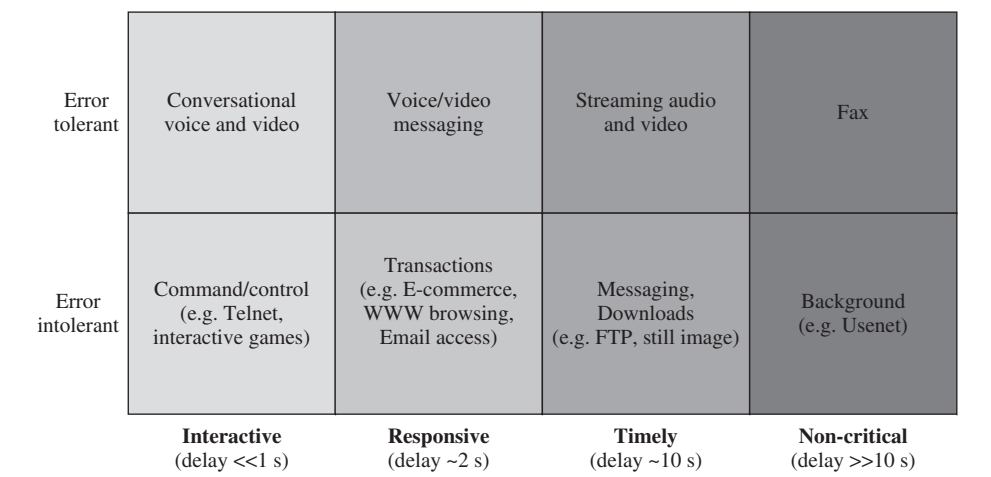


Figure 1.24 Model for user-centric QoS categories (Source: ITU 2001 [18]. Reproduced with permission of ITU.)

This mapping is summarised in Figure 1.24, which provides a recommended model for end-user QoS categories, where the four areas of delay are given names chosen to illustrate the type of user interaction involved.

1.13.8 Network Performance

Network performance (NP) contributes towards QoS as experienced by the user/customer. Network performance may or may not be on an end-to-end basis. For example, access performance is usually separated from the core network performance in the operations of a single IP

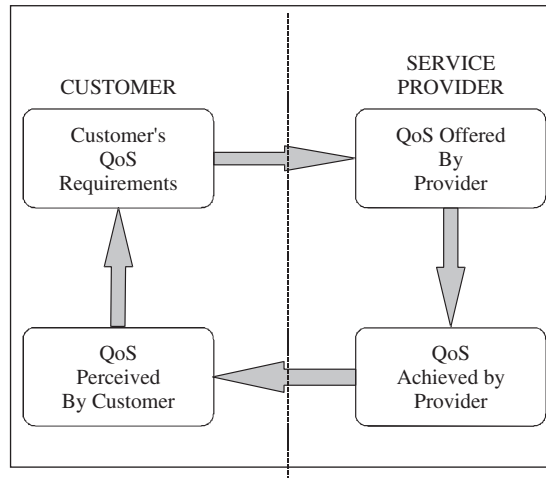


Figure 1.25 The four viewpoints of QoS (Source: ITU 2001 [17]. Reproduced with permission of ITU.)

network, while Internet performance often reflects the combined NP of several autonomous networks.

There are four viewpoints of QoS defined by the ITU-T G.1000 recommendation, corresponding with different perspectives, as shown in Figure 1.25:

- customer QoS requirements;
- service provider offerings of QoS (or planned/targeted QoS);
- QoS achieved or delivered;
- customer survey ratings of QoS.

Among these four viewpoints, the customer's QoS requirements may be considered as the logical starting point. A set of customer's QoS requirements may be treated in isolation as far as customer's concerns are captured. This requirement is an input to the service provider for the determination of the QoS to be offered or planned.

1.13.9 QoS and NP for Satellite Networking

The definitions of QoS given by the ITU-T are based on a user-centric approach, but these may not reflect well on the QoS and NP related to networking. Therefore it is useful to employ the layering approach to define QoS and NP parameters related to networks called network centric approach (see Figure 1.26).

The network centric approach enables us to quantify the QoS and NP parameters without the uncertainty of terminal performance, higher layer protocol functions and user factors. Typical parameters are:

- at analogue transmission level: signal to noise power ratio (S/N);
- at digital transmission level: bit error ratio (BER), propagation delay and delay variation;

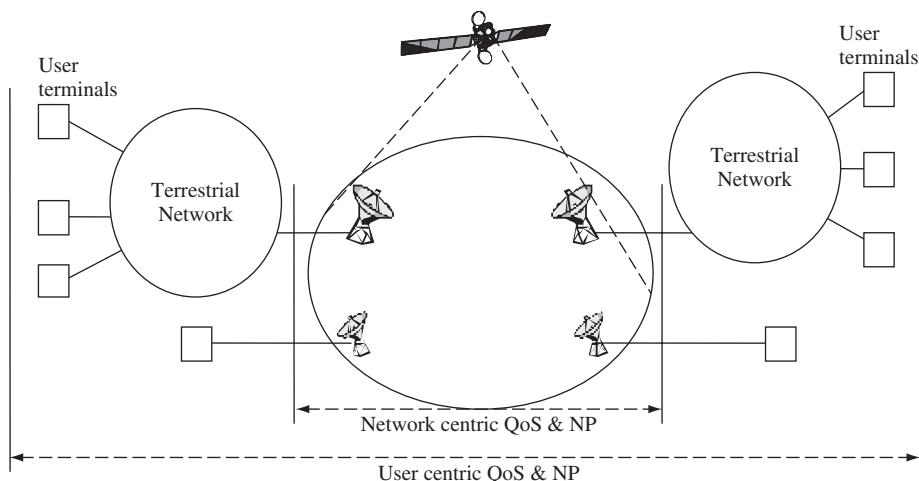


Figure 1.26 User- and network-centric views of QoS and NP concepts

- at packet level: packet propagation delay and packet delay variation, packet error ratio, packet loss ratio and network throughput.

1.14 Digital Video Broadcasting (DVB)

Digital video broadcasting (DVB) technology allows broadcasting of ‘data containers’, in which all kinds of digital data can be transmitted. It simply delivers compressed images, sound or data to the receiver within these ‘containers’. No restriction exists as to the kind of information in the data containers. The DVB ‘service information’ acts like a header to the container, ensuring that the receiver knows what it needs to decode.

A key difference of the DVB approach compared to other data broadcasting systems is that the different data elements within the container can carry independent timing information. This allows, for example, audio information to be synchronised with video information in the receiver, even if the video and audio information does not arrive at the receiver at exactly the same time.

This facility is, of course, essential for the transmission of conventional television programmes. The DVB approach provides a good deal of flexibility. For example, a 38 Mbit/s data container could hold eight standard definition television (SDTV) programmes, four enhanced definition television (EDTV) programmes or one high-definition television (HDTV) programme, all with associated multi-channel audio and ancillary data services.

Alternatively, a mix of SDTV and EDTV programmes could be provided or even multimedia data containing little or no video information. The content of the container can be modified to reflect changes in the service offer over time (e.g. migration to a widescreen presentation format).

At present, the majority of DVB satellite transmissions convey multiple SDTV programmes and associated audio and data but more and more HDTV have been introduced. DVB is also useful for data broadcasting services (e.g. access to the World Wide Web).

1.14.1 The DVB Standards

Digital video broadcasting (DVB) is a term that is generally used to describe digital television and data broadcasting services that comply with the DVB 'standard'.

In fact, there is no single DVB standard, but rather a collection of standards, technical recommendations and guidelines. These were developed by the Project on Digital Video Broadcasting, usually referred to as the 'DVB Project'.

The DVB Project was initiated in 1993 in liaison with the European Broadcasting Union (EBU), the European Telecommunications Standards Institute (ETSI) and the European Committee for Electrotechnical Standardisation (CENELEC). As opposed to traditional governmental agency standards activities round the world, the DVB Project is market-driven and consequently works on commercial terms, to tight deadlines and realistic requirements, always with an eye toward promoting its technologies through achieving economies of scale. Though based in Europe, the DVB Project is international. DVB specifications concern:

- source coding of audio, data and video signals;
- channel coding;
- transmitting DVB signals over terrestrial and satellite communications paths;
- scrambling and conditional access;
- the general aspects of digital broadcasting;
- software platforms in user terminals;
- user interfaces supporting access to DVB services;
- the return channel, as from a user back to an information or programme source to support interactive services.

The DVB specifications are interrelated with other recognised specifications. DVB source coding of audio-visual information as well as multiplexing is based on the standards evolved by the Moving Picture Experts Group (MPEG), a joint effort of the International Organisation for Standards (ISO) and the International Electrotechnical Commission (IEC). The principal advantage of MPEG compared to other audio and audio coding formats is that the sophisticated compression techniques used make MPEG files far smaller for the same quality.

DVB contains a family of standard: DVB-S for satellite, TVB-T for terrestrial, DVB-H for handheld terminals, DVB-C for cable and so on.

1.14.2 Transmission System

The transmission system consists of the functional block of equipment to transport the base-band TV signals in the format of the MPEG-2 transport stream over the satellite channel. The transmission system carries out the following block functions of processes on the data stream as shown in Figure 1.27:

- transport multiplex adaptation and randomization for energy dispersal;
- outer coding (i.e. Reed–Solomon);
- convolutional interleaving;
- inner coding (i.e. punctured convolutional code);

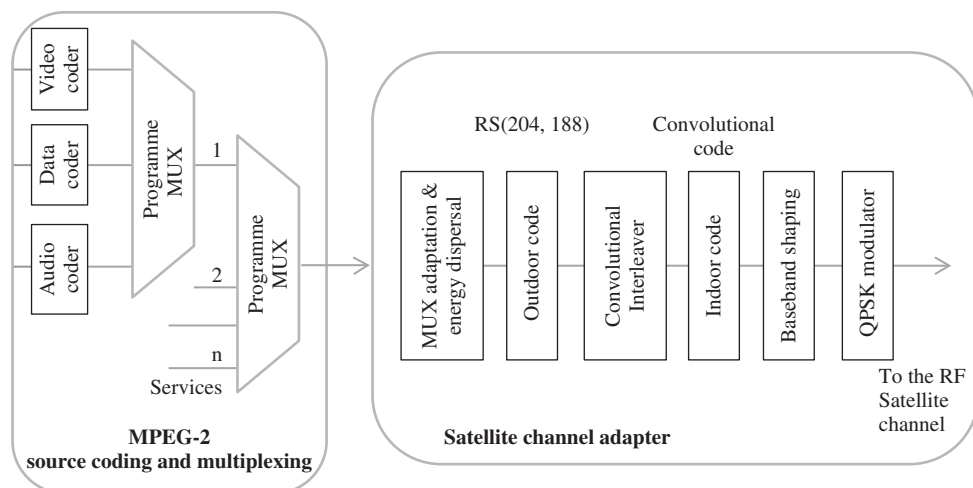


Figure 1.27 Functional blocks of the transmission system (Source: ETSI 1997 [8]. Reproduced with permission of ETSI.)

- baseband shaping for modulation;
- modulation.

MPEG-2 source coding and multiplexing are considered to be outside the scope of the DVB-S standard and will not be discussed further here.

As digital satellite TV services are particularly affected by power limitations, the DVT-S is designed to overcome noise and interference rather than to achieve spectrum efficiency. To achieve a high power efficiency without excessively penalizing the spectrum efficiency, the DVB-S uses QPSK modulation and the concatenation of convolutional and RS codes. The convolutional code can be configured flexibly, allowing the optimization of the system performance for a given satellite transponder bandwidth.

It is optimized for single carrier per transponder time division multiplex (TDM), but can also be used for multi-carrier frequency division multiplex (FDM) applications.

The DVB-S is directly compatible with MPEG-2 coded TV signals (defined by ISO/IEC DIS 13818-1). The modem transmission frame is synchronous with the MPEG-2 multiplex transport packets. If the received signal is above C/N and C/I threshold, the Forward Error Correction (FEC) technique can provide a ‘quasi error free’ (QEF) quality target. The QEF means that Bit Error Ratio (BER) less than 10^{-10} to 10^{-11} at the input of the MPEG-2 demultiplexer.

1.14.3 Adaptation to Satellite Transponder Characteristics

Transmissions of digital multi-programme TV services will use satellites in both the fixed satellite service (FSS) and the broadcast satellite service (BSS) bands. The choice of transponder bandwidth is a function of the satellite used and the data rates required by the service. Table 1.3 gives the specification of the DVB-S interfaces.

Loading of the sequence ‘100101010000000’ into the PRBS registers, as indicated in Figure 1.28, is initiated at the start of every eight transport packets. To provide an initialization signal for the descrambler, the MPEG-2 sync byte of the first transport packet in a group of eight packets is bit-wise inverted from 47HEX to B8HEX. This process is referred to as the ‘transport multiplex adaptation’.

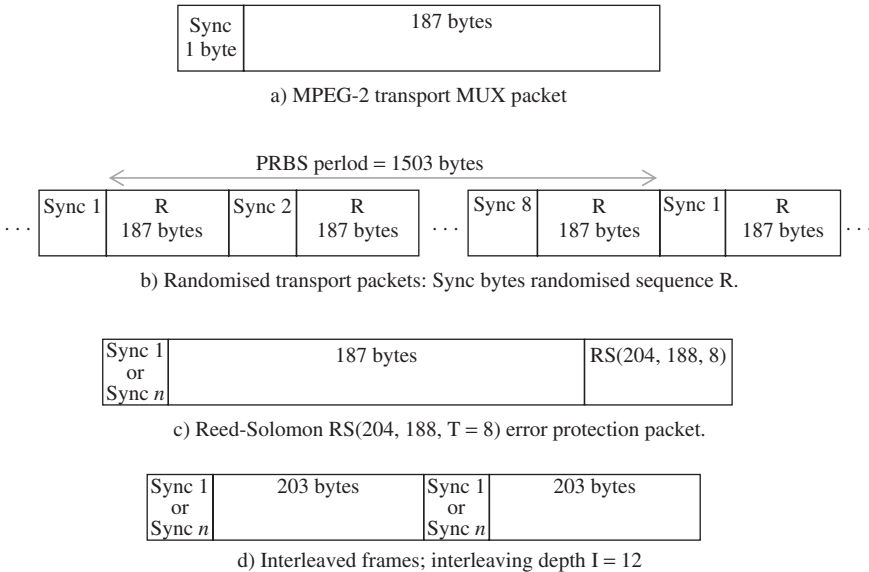
The first bit at the output of the PRBS generator is applied to the first bit (i.e. MSB) of the first byte following the inverted MPEG-2 sync byte (i.e. B8HEX). To aid other synchronization functions, during the MPEG-2 sync bytes of the subsequent seven transport packets, the PRBS generation continues, but its output is disabled, leaving these bytes unrandomised. Thus, the period of the PRBS sequence is 1503 bytes.

The randomisation process is also active when the modulator input bit-stream is non-existent, or when it is noncompliant with the MPEG-2 transport stream format (i.e. 1 sync byte + 187 packet bytes). This is to avoid the emission of an unmodulated carrier from the modulator.

1.14.5 Reed–Solomon (RS) Outer Coding, Interleaving and Framing

The framing organisation is based on the input packet structure as shown in Figure 1.29a.

Reed-Solomon RS(204, 188, $T = 8$) shortened code, from the original RS(255, 239, $T = 8$) code, is applied to each randomized transport packet (188 bytes) of Figure 1.29b to generate an



Sync 1 = not randomised complemented sync byte
Sync n = not randomised sync byte, $n = 2, 3, \dots, 8$

Figure 1.29 Framing structure (Source: ETSI 1997 [8]. Reproduced with permission of ETSI.)

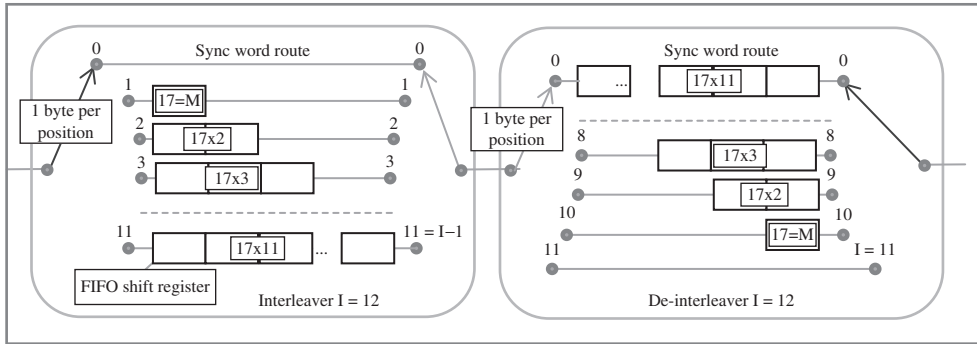


Figure 1.30 Conceptual diagram of the convolutional interleaver and de-interleaver (Source: ETSI 1997 [8]. Reproduced with permission of ETSI.)

error protected packet (see Figure 1.29c). Reed–Solomon is applied to the packet sync-byte, either non-inverted (i.e. 47HEX) or inverted (i.e. B8HEX).

Code generator polynomial: $g(x) = (x + \lambda^0)(x + \lambda^1)(x + \lambda^2) \dots (x + \lambda^{15})$, where $\lambda = 02_{\text{HEX}}$.

Field generator polynomial: $p(x) = x^8 + x^4 + x^3 + x^2 + 1$.

The shortened Reed–Solomon code is implemented by adding 51 bytes, all set to zero, before the information bytes at the input of a (255 239) encoder. After the RS coding procedure these null bytes shall be discarded.

Following the conceptual scheme of Figure 1.30, convolutional interleaving with depth $I = 12$ is applied to the error protected packets (see Figure 1.29c). This produces an interleaved frame (see Figure 1.29d).

The interleaved frame consists of overlapping error protected packets and is delimited by inverted or non-inverted MPEG-2 sync-bytes (preserving the periodicity of 204 bytes).

The interleaver consists of $I = 12$ branches, cyclically connected to the input byte-stream by the input switch. Each branch is a first-in-first-out (FIFO) shift register, with depth (Mj) cells (where $M = 17 = N/I$; $N = 204 =$ error protected frame length, $I = 12 =$ interleaving depth, $j =$ branch index). The cells of the FIFO shall contain 1 byte, and the input and output switches shall be synchronized. For synchronization purposes, the sync bytes and the inverted sync bytes is always routed in the branch ‘0’ of the interleaver (corresponding to a null delay).

The de-interleaver is similar, in principle, to the interleaver, but the branch indexes are reversed (i.e. $j = 0$ corresponds to the largest delay). The de-interleaver synchronization is carried out by routing the first recognized sync byte in the ‘0’ branch.

1.14.6 Inner Convolutional Coding

The DVB-S allows for a range of punctured convolutional codes, based on a rate 1/2 convolutional code with constraint length $K = 7$. This will allow selection of the most appropriate level of error correction for a given service data rate. It allows convolutional coding with code

Table 1.4 Punctured code definition

Original code			Code rates									
			1/2		2/3		3/4		5/6		7/8	
K	G1 (X)	G2 (Y)	P	d _{free}	P	d _{free}	P	d _{free}	P	d _{free}	P	d _{free}
7	171 _{oct}	133 _{oct}	X:1 Y:1	10	X:10 Y:11	6	X:101 Y:110	5	X:10101 Y:11010	4	X:1000101 Y:1111010	3
			I=X1 Q=Y1		I=X1Y2Y3 Q=Y1X3Y4		I=X1Y2 Q=Y1X3		I=X1Y2Y4 Q=Y1X3X5		I=X1Y2Y4Y6 Q=Y1X3X5X7	

Note: 1 = transmitted bit; 0 = non transmitted bit
Source: ETSI 1997 [8]. Reproduced with permission of ETSI.

rates of 1/2, 2/3, 3/4, 5/6 and 7/8. Table 1.4 gives the punctured convolutional code (also see Figure 1.31 in the next section).

1.14.7 Baseband Shaping and Modulation

The DVB-S employs conventional Gray-coded QPSK modulation with absolute mapping (no differential coding). Figure 1.31 shows the bit mapping in the signal space. Prior to modulation, the I and Q signals (mathematically represented by a succession of Dirac delta functions spaced

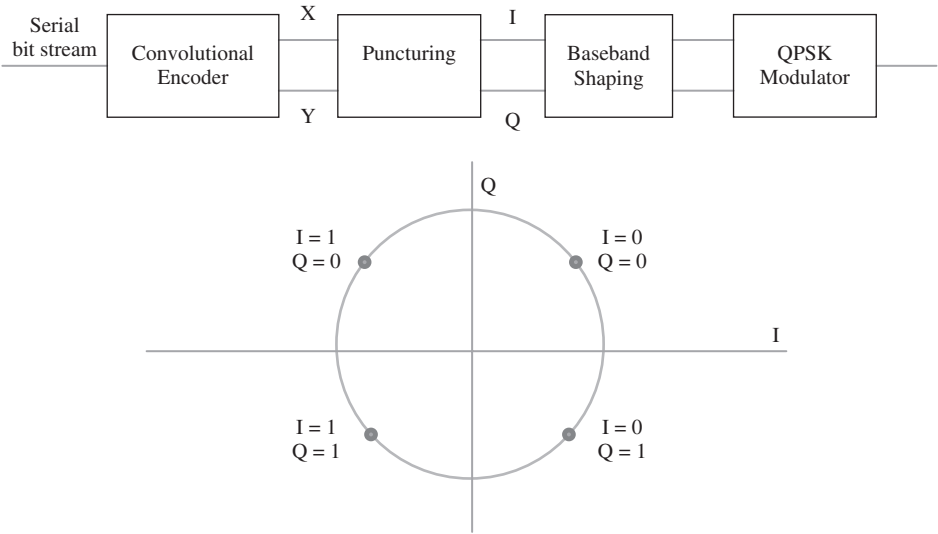


Figure 1.31 QPSK constellation (Source: ETSI 1997 [8]. Reproduced with permission of ETSI.)

Table 1.5 IF loop performance of the system

Inner code rate	Required E_b/N_0 for BER = 2×10^{-4} after Viterbi QEF after Reed–Solomom
1/2	4.5
2/3	5.0
3/4	5.5
5/6	6.0
7/8	6.4

Note 1: The figures of E_b/N_0 refer to the useful bit-rate before RS coding and include a modem implementation margin of 0.8 dB and the noise bandwidth increase due to the outer code ($10 \log 188/204 = 0.36$ dB).

Note 2: Quasi-error-free (QEF) means less than one uncorrected error event per hour, corresponding to BER = 10^{-10} to 10^{-11} at the input of the MPEG-2 demultiplexer.

Source: ETSI 1997 [8]. Reproduced with permission of ETSI.

by the symbol duration $T_s = 1/R_s$, with appropriate sign) is square root raised cosine filtered. The value of 0.35 is selected as the roll-off factor.

The following expression is a theoretical function definition for the baseband square root raised cosine filter:

$$H(f) = \begin{cases} 1, & \text{for } |f| < f_N(1 - \alpha) \\ \sqrt{\frac{1}{2} + \frac{1}{2} \sin \frac{\pi}{2f_N} \left[\frac{f_N - |f|}{\alpha} \right]}, & \text{for } f_N(1 - \alpha) \leq |f| \leq f_N(1 + \alpha) \\ 0, & \text{for } |f| > f_N(1 + \alpha) \end{cases} \quad (1.12)$$

Where $f_N = \frac{1}{2T_s} = \frac{R_s}{2}$ is the Nyquist frequency and $\alpha = 0.35$ is the roll-off factor.

1.14.8 Error Performance Requirements

Table 1.5 gives the modem BER versus E_b/N_0 performance requirements for the system IF loop.

1.15 DVB-S Satellite Delivery

One of the earliest standards developed by the DVB project and formulated by ETSI was for digital video broadcasting via satellite (usually referred to as the ‘DVB-S standard’). Specifications also exist for the retransmission of DVB signals via cable networks and satellite master antenna television (SMATV) distribution networks.

The techniques used for DVB via satellite are classical in the sense that they have been used for many years to provide point-to-point and point-to-multipoint satellite data links in 'professional' applications. The key contribution of the DVB project in this respect has been the development of highly integrated and low-cost chip sets that adapt the DVB baseband signal to the satellite channel. Data transmissions via satellite are very robust, offering a maximum bit error rate in the order of 10^{-11} .

In satellite applications, the maximum data rate for a data container is typically about 38 Mbit/s. This container can be accommodated in a single 33 MHz satellite transponder. It provides sufficient capacity to deliver, for example, four to eight standard television programmes, 150 radio channels, 550 ISDN channels or any combination of these services. This represents a significant improvement over conventional analogue satellite transmission, where the same transponder is typically used to accommodate a single television programme with far less operational flexibility.

A single modern high-power broadcasting satellite typically provides at least twenty 33 MHz transponders, allowing delivery of about 760 Mbit/s of data to large numbers of users equipped with small (around 60 cm) satellite dishes. The new generation of broadband satellite can now capable of delivering a total capacity of over Gbit/s of data services.

A simple generic model of a digital satellite transmission channel comprises several basic building blocks, which include baseband processing and channel adaptation in the transmitter and the complementary functions in the receiver. Central to the model is, of course, the satellite transmission channel. Channel adaptation would most likely be done at the transmit satellite earth station, while the baseband processing would be performed at a point close to the programme source.

1.15.1 MPEG-2 Baseband Processing

MPEG is a group of experts drawn from industry who contribute to the development of common standards through an ITU-T and ISO/IEC joint committee. The established MPEG-2 standard was adopted in DVB for the source coding of audio and video information and for multiplexing a number of source data streams and ancillary information into a single data stream suitable for transmission. Therefore, many of the parameters, fields and syntax used in DVB baseband processing are specified in the relevant MPEG-2 standards. The MPEG-2 standards are generic and very wide in scope. Some of the parameters and fields of MPEG-2 are not used in DVB.

The processing function deals with a number of programme sources. Each programme source comprises any mixture of raw data and uncompressed video and audio, where the data can be, for example, teletext and/or subtitling information and graphical information such as logos.

Each of the video, audio and programme-related data is called an elementary stream (ES). It is encoded and formatted into a packetised elementary stream (PES). Thus each PES is a digitally encoded component of a programme.

The simplest type of service is a radio programme, which would consist of a single audio elementary stream. A traditional television broadcast would comprise three elementary streams: one carrying coded video, one carrying coded stereo audio and one carrying teletext.

1.15.2 Transport Stream (TS)

Following packetisation, the various elementary streams of a programme are multiplexed with packetised elementary streams from other programmes to form a transport stream (TS). Each of the packetised elementary streams can carry timing information, or 'time stamps', to ensure that related elementary streams, for example, video and audio, are replayed in synchronism in the decoder. Programmes can each have a different reference clock, or can share a common clock. Samples of each 'programme clock', called programme clock references (PCRs), are inserted into the transport stream to enable the decoder to synchronise its clock to that in the multiplexer. Once synchronised, the decoder can correctly interpret the time stamps and can determine the appropriate time to decode and present the associated information to the user.

Additional data is inserted into the transport stream, which includes programme specific information (PSI), service information (SI), conditional access (CA) data and private data. Private data is a data stream whose content is not specified by MPEG.

The transport stream is a single data stream that is suitable for transmission or storage. It may be of fixed or variable data rate and may contain fixed or variable data rate elementary streams. There is no form of error protection within the multiplex. Error protection is implemented within the satellite channel adaptor.

1.15.3 Service Objectives

The DVB-S system is designed to provide so-called 'quasi error free' (QEF) quality. This means less than one uncorrected error event per transmission hour, corresponding to a bit error rate (BER) of between 10^{-10} and 10^{-11} at the input of the MPEG-2 demultiplexer (i.e. after all error correction decoding). This quality is necessary to ensure that the MPEG-2 decoders can reliably reconstruct the video and audio information.

This quality target translates to a minimum carrier-to-noise (C/N) ratio requirement for the satellite link, which in turn determines the requirements for the transmit earth station and the user's satellite reception equipment for a given satellite broadcasting network. The requirement is actually expressed in E_b/N_0 (energy per bit to noise density ratio), rather than (C/N), so that it is independent of the transmission rate.

The DVB-S standard specifies the E_b/N_0 values at which QEF quality must be achieved when the output of the modulator is directly connected to the input of the demodulator (i.e. in an 'IF loop'). An allowance is made for practical implementation of the modulator and demodulator functions and for the small degradation introduced by the satellite channel. The values range from 4.5 dB for rate 1/2 convolutional coding to 6.4 dB for rate 7/8 convolutional coding.

The inner code rate can be varied to increase or decrease the degree of error protection for the satellite link at the expense of capacity. The reduction or increase in capacity associated with a change in the code rate and the related increase or reduction in the E_b/N_0 requirement. The latter is also expressed as an equivalent increase or reduction in the diameter of the receive antenna (the size of user's satellite dish), all other link parameters remaining unchanged.

1.15.4 Satellite Channel Adaptation

The DVB-S standard is intended for direct-to-home (DTH) services to consumer integrated receiver decoders (IRD), as well as for reception via collective antenna systems (satellite

master antenna television; SMATV) and at cable television head-end stations. It can support the use of different satellite transponder bandwidths, although a bandwidth of 33 MHz is commonly used. All service components ('programmes') are time division multiplexed (TDM) into a single MPEG-2 transport stream, which is then transmitted on a single digital carrier.

The modulation is classical quadrature phase shift keying (QPSK). A concatenated error protection strategy is employed based on a convolutional 'inner' code and a shortened Reed–Solomon (RS) 'outer' code. Flexibility is provided so that transmission capacity can be traded off against increased error protection by varying the rate of the convolutional code. Satellite links can therefore be made more robust, at the expense of reduced throughput per satellite transponder (i.e. fewer DVB services).

The standard specifies the characteristics of the digitally modulated signal to ensure compatibility between equipment developed by different manufacturers. The processing at the receiver is, to a certain extent, left open to allow manufacturers to develop their own proprietary solutions. It also defines service quality targets and identifies the global performance requirements and features of the system that are necessary to meet these targets.

1.15.5 DVB Return Channel over Satellite (DVB-RCS)

The principal elements of a DVB return channel over satellite (DVB-RCS) system are the hub station and user satellite terminals. The hub station controls the terminals over the forward (also called outbound link), and the terminals share the return (also called inbound link). The hub station continuously transmits the forward link in time division multiplex (TDM). The terminals transmit as needed, sharing the return channel resources using multi-frequency time division multiple access (MF-TDMA). The DVB-RCS system supports communications on channels in two directions:

- forward channel, from the hub station to many terminals;
- return channels, from the terminals to the hub station.

The forward channel is said to provide 'point-to-multipoint' service, because it is sent by a station at a single point to stations at many different points. It is identical to a DVB-S broadcast channel and has a single carrier, which may take up the entire bandwidth of a transponder (bandwidth-limited) or use the available transponder power (power limited). Communications to the terminals share the channel by using different slots in the TDM carrier.

The terminals share the return channel capacity of one or more satellite transponders by transmitting in bursts, using MF-TDMA. In a system, this means that there is a set of return channel carrier frequencies, each of which is divided into time slots which can be assigned to terminals, which permits many terminals to transmit simultaneously to the hub. The return channel can serve many purposes and consequently offers choices of some channel parameters. A terminal can change frequency, bit rate, FEC rate, burst length or all of these parameters, from burst to burst. Slots in the return channel are dynamically allocated.

The uplink and downlink transmission times of the signaling propagation between the hub and the satellite are very nearly fixed. However, the terminals are at different points, so the signal transit times between them and the satellite vary. On the forward channel, this variation is unimportant. Just as satellite TV sets successfully receive signals whenever they arrive, the terminals receive downlink signals without regard to small differences in their times of arrival.

However, on the uplink, in the return direction from the terminals to the hub, small differences in transit time can disrupt transmission. This is because the terminals transmit in bursts that share a common return channel by being spaced from each other in time. For instance, a burst from one terminal might be late because it takes longer to reach the satellite than a burst sent by another terminal. A burst that is earlier or later than it should be can collide with the bursts sent by the terminals using the neighbouring TDMA slots.

The difference in transmission times to terminals throughout the footprint of a satellite might be compensated for by using time slots that are considerably longer than the bursts transmitted by the terminals, so both before and after a burst there is a guard time sufficiently long to prevent collisions with the bursts in neighbouring slots in the TDMA frame. The one-way delay time between a hub and a terminal varies from 250 to 290 ms, depending on the geographical location of the terminal with respect to the hub. So the time differential, T , might be as large as 40 ms. So most TDMA satellite systems minimise guard time by incorporating various means of timing adjustment to compensate for satellite path differences. DVB-RCS has two built-in methods of pre-compensating the burst transmission time of each terminal:

- Each terminal ‘knows’ its local GPS coordinates and therefore can calculate its own burst transmission time.
- The hub monitors the arrival times of bursts and can send correction data to terminals if need be.

1.15.6 TCP/IP over DVB

DVB-RCS uses the MPEG-2 digital wrappers, in which ‘protocol-independent’ client traffic is enclosed within the payloads of a stream of 188-byte packets. The MPEG-2 digital wrapper offers a 182-byte payload and has a 6-byte header. The sequence for transmission of Internet TCP/IP traffic includes:

- The TCP/IP message arrives and is subjected to TCP optimisation.
- The IP packets are divided into smaller pieces and put into data sections with 96-bit digital storage medium – command and control (DSM-CC) headers.
- The DSM-CC data sections are further divided into 188-byte MPEG2-TS packets in the baseband processing.
- The MPEG2-TS packets then are subjected to channel coding for satellite transmissions.

1.16 DVB Satellite – Second Generation (DVB-S2)

The DVB-S standard uses QPSK modulation and concatenated convolutional and Reed-Solomon channel coding. It has been adopted by most satellite operators worldwide for television and data broadcasting services. Digital satellite transmission technology has evolved significantly in several areas, since the first publication of the DVB-S standard in 1994. Without going too many details of the DVB-S2 standard, this section provides a brief summary of the DVB-S2 technology novelty, transmission system architecture and performance.

1.16.1 Technology Novelty in the DVB-S2

The DVB-S2 tried to make use of the new development of the technology and future applications of broadband satellite applications. The main features can be summarised as the following:

- New channel coding schemes to achieve a capacity gain in the order of 30%;
- Variable coding and modulation (VCM) is used to provide different levels of error protection to different service components (e.g. SDTV and HDTV, audio, multimedia).
- In the case of interactive and point-to-point applications, the VCM functionality is combined with the use of return channels, to achieve Adaptive Coding and Modulation (ACM). This technique provides dynamic link adaptation to propagation conditions, targeting each individual receiving terminal. ACM systems promise satellite capacity gains of up to 100–200%. Such gains are achieved by informing the satellite up-link station of the channel condition (e.g. $C/N+I$) of each receiving terminal via the satellite or terrestrial return channels.
- Extended flexibility to cope with other input data formats (such as multiple Transport Streams, or generic data formats) in addition to the single MPEG transport stream (MPEG-TS) in DVB-S without significant complexity increase.

The DVB-S2 makes use of the new technology in the following functions:

- Stream adapter, suitable for operation with single and multiple input streams of various formats (packetised or continuous);
- Forward error correction (FEC) based on low density parity check (LDPC) codes concatenated with BCH codes, allowing quasi error free (QEF) operation at about 0.7–1.0 dB from the Shannon limit;
- A wide range of code rates (from 1/4 up to 9/10); four constellations, ranging in spectrum efficiency from 2 bit/s/Hz to 5 bit/s/Hz, optimized for operation over non-linear transponders;
- Three spectrum shapes with roll-off factors 0.35, 0.25 and 0.20;
- Adaptive coding and modulation (ACM) functionality, optimizing channel coding and modulation on a frame-by-frame basis.

The DVB-S2 has also been design to support a wide range of broadband satellite applications including:

- Broadcast services (BS) digital multi-programme television (TV)/high definition television (HDTV) – for primary and secondary distribution in the fixed satellite service (FSS) and the broadcast satellite service (BSS) bands. The DVB-S2 provides direct-to-home (DTH) services for consumer integrated receiver decoder (IRD), as well as collective antenna systems (satellite master antenna television – SMATV) and cable television head-end stations (possibly with remodulation. It has two modes: non-backwards compatible broadcast services (NBC-BS) for exploitation of full benefit of the DVB-S2 but not compatible with DVB-S; and backwards-compatible broadcast services (BC-BS) is backwards-compatible DVB-S to give time for migration from DVB-S to DVB-S2.

- Interactive services (IS) interactive data services including Internet access – for providing interactive services to consumer Integrated receiver decoders (IRD) and to personal computers, where DVB-S2's forward path supersedes the current DVB-S for interactive systems. The return path can be implemented using various DVB interactive systems, such as DVB-RCS (EN 301 790), DVB-RCP (ETS 300 801), DVB-RCG (EN 301 195) and DVB-RCC (ES 200 800).
- Digital TV contribution and satellite news gathering (DTVC/DSNG) – for temporary and occasional transmission with short notice of television or sound for broadcasting purposes, using portable or transportable uplink earth stations. Digital television contribution applications by satellite consist of point-to-point or point-to-multipoint transmissions, connecting fixed or transportable uplink and receiving stations. They are not intended for reception by the general public.
- Data content distribution/trunking and other professional applications (PS) – for point-to-point or point-to-multipoint, including interactive services to professional head-ends, which re-distribute services over other media. Services may be transported in (single or multiple) generic stream format.

Digital transmissions via satellite are affected by power and bandwidth limitations. DVB-S2 tries to overcome these limits by making use of transmission modes (FEC coding and modulations), giving different trade-offs between power and spectrum efficiency.

For some specific applications (e.g. broadcasting), modulation techniques such as QPSK and 8PSK, with their quasi-constant envelope, are appropriate for operation with saturated satellite power amplifiers (in single carrier per transponder configuration). When higher power margins are available, spectrum efficiency can be further increased to reduce bit delivery cost. In these cases also 16APSK and 32APSK can operate in single carrier mode close to the satellite HPA saturation by pre-distortion techniques.

DVB-S2 is compatible with Moving Pictures Experts Group (MPEG-2 and MPEG-4) coded TV services (ISO/IEC 13818-1), with a transport stream packet multiplex. All service components are time division multiplexed (TDM) on a single digital carrier.

1.16.2 Transmission System Architecture

The DVB-S2 system consists of a number of functional block of equipment performing the adaptation of the baseband digital signals, from the output of a single (or multiple) MPEG transport stream multiplexer(s) (ISO/IEC 13818-1), or from the output of a single (or multiple) generic data source(s), to the satellite channel characteristics.

Data services may be transported in transport stream format according to EN 301 192 (e.g. using multi-protocol encapsulation), or generic stream format.

The DVB-S2 provides a 'quasi error free' (QEF) quality target 'less than one uncorrected error-event per transmission hour at the level of a 5 Mbit/s single TV service decoder', approximately corresponding to a transport stream packet error ratio (PER) of less than 10^{-7} before de-multiplexer.

Figure 1.32 illustrates these function blocks which are explained in the following:

- *Mode adaptation* is application dependent. It provides the following function blocks
 - input stream interfacing;
 - input stream synchronization (optional);
 - null-packet deletion (for ACM and transport stream input format only);

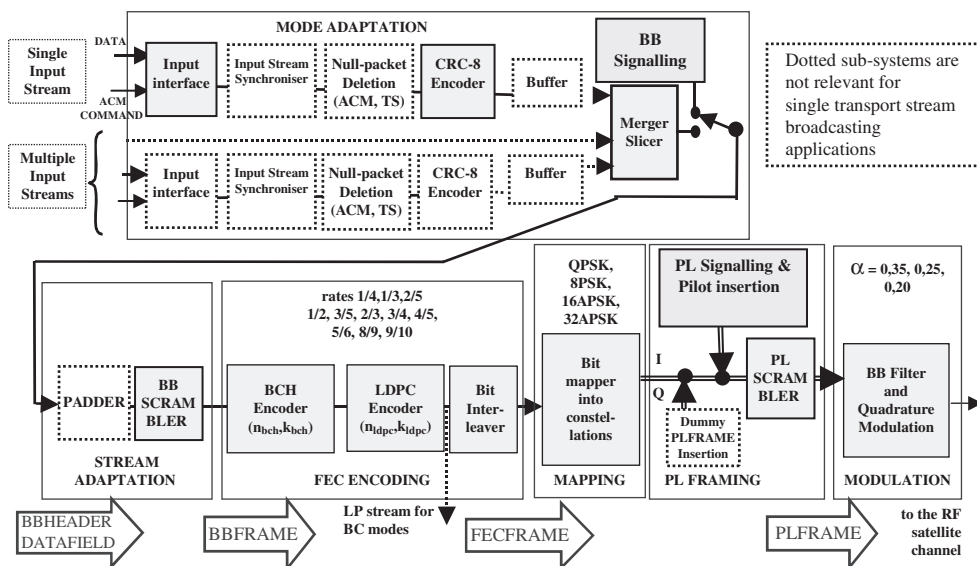


Figure 1.32 Functional block diagram of the DVB-S2 system (Source: ETSI 2013 [12]. Reproduced with permission of ETSI.)

- CRC-8 coding for error detection at packet level in the receiver (for packetised input streams only);
- merging of input streams (for multiple input stream modes only) and slicing into *data fields*;
- a base-band header shall be appended in front of the data field, to notify the receiver of the input stream format and mode adaptation type. To be noted that the MPEG multiplex transport packets may be asynchronously mapped to the base-band frames.
- *Stream adaptation* has two functions:
 - to provide padding to complete a base-band frame;
 - base-band scrambling.
- *Forward error correction (FEC) encoding* is carried out by the concatenation of coding functions and one interleaving functions:
 - BCH outer codes;
 - low density parity check (LDPC) inner codes (rates 1/4, 1/3, 2/5, 1/2, 3/5, 2/3, 3/4, 4/5, 5/6, 8/9, 9/10);
 - bit interleaving functions is applied to FEC coded bits for 8PSK, 16APSK and 32APSK.
- *Mapping function* maps the bit stream of the FEC into QPSK, 8PSK, 16APSK and 32APSK constellations, depending on the application area. Gray mapping of constellations is used for QPSK and 8PSK.
- *Physical layer framing (PLFRAME)* is used for synchronous with the FEC frames, to provide dummy PLFRAME insertion with three functions:
 - physical layer (PL) signalling, pilot symbols insertion (optional);
 - physical layer scrambling for energy dispersal;

- dummy PLFRAMEs are transmitted when no useful data is ready to be sent on the channel.
- *Base-band filtering and quadrature modulation* is to shape the signal spectrum (squared-root raised cosine, roll-off factors 0,35 or 0,25 or 0,20) and to generate the RF signal.

1.16.3 Error Performance

The error performance is described to meet the requirements of quasi error free (QEF) over AWGN (E_s = average energy per transmitted symbol). Table 1.6 illustrates the error

Table 1.6 Es/No performance at quasi error free PER = 10^{-7} (AWGN channel)

Mode		Spectral efficiency	Ideal E_s/N_0 (dB) for FECFRAME length = 64800
QPSK	QPSK 1/4	0.490243	-2.36
	QPSK 1/3	0.656448	-1.24
	QPSK 2/5	0.789412	-0.30
	QPSK 1/2	0.988858	1.00
	QPSK 3/5	1.188304	2.23
	QPSK 2/3	1.322253	3.10
	QPSK 3/4	1.487473	4.03
	QPSK 4/5	1.587196	4.68
	QPSK 5/6	1.654663	5.18
	QPSK 8/9	1.766451	6.20
8PSK	QPSK 9/10	1.788612	6.42
	8PSK 3/5	1.779991	5.50
	8PSK 2/3	1.980636	6.62
	8PSK 3/4	2.228124	7.91
	8PSK 5/6	2.478562	9.35
	8PSK 8/9	2.646012	10.69
16APSK	8PSK 9/10	2.679207	10.98
	16APSK 2/3	2.637201	8.97
	16APSK 3/4	2.966728	10.21
	16APSK 4/5	3.165623	11.03
	16APSK 5/6	3.300184	11.61
	16APSK 8/9	3.523143	12.89
32APSK	16APSK 9/10	3.567342	13.13
	32APSK 3/4	3.703295	12.72
	32APSK 4/5	3.951571	13.64
	32APSK 5/6	4.119540	14.28
	32APSK 8/9	4.397854	15.69
	32APSK 9/10	4.453027	16.01

Note: Given the system spectral efficiency η_{tot} the ratio between the energy per information bit and single side noise power spectral density $E_b/N_0 = E_s/N_0 - 10 \log(\eta_{\text{tot}})$

Source: ETSI 2013 [12]. Reproduced with permission of ETSI.

performance provided in the DVB-S2 standard, where ideal E_s/N_0 (dB) is the figure achieved by computer simulation, 50 LDPC fixed point decoding iterations, perfect carrier and synchronization recovery, no phase noise, AWGN channel.

The standard also suggest that for short forward error correction frames (FECFRAME) an additional degradation of 0.2–0.3 dB has to be taken into account; for calculating link budgets, specific satellite channel impairments should be taken into account. The packet error ratio (PER) is defined as the ratio between the useful transport stream packets (188 bytes) correctly received and affected by errors, after forward error correction. Spectral efficiencies (per unit symbol rate) are computed for normal forward error correction frame (FECFRAME) length and no pilots.

1.17 DVB Satellite Services to Handheld Devices (DVB-SH)

The DVB-SH is an ETSI standard which specifies a transmission system for hybrid satellite and terrestrial digital television broadcasting to mobile or handheld terminals. It is derived from the DVB-T and DVB-H system specification, respectively designed for digital television terrestrial broadcasting towards fixed and mobile terminals and DVB-S2, designed for digital satellite broadcasting towards fixed terminals. Information of this section is based on the ETSI EN 302 583 V1.1.2 (2010-02): Digital video broadcasting (DVB), framing structure, channel coding and modulation system for satellite services to handheld devices (SH) below 3 GHz.

The DVB-SH standard specifies a system of satellite services to handheld devices using frequencies below 3 GHz.

This relies on a hybrid satellite/terrestrial infrastructure. The signals are broadcast to mobile terminals on two paths:

- A direct path from a broadcast station to the terminals via the satellite.
- An indirect path from a broadcast station to terminals via terrestrial repeaters that form the complementary ground component (CGC) to the satellite. The CGC can be fed through satellite and/or terrestrial distribution networks.

The DVB-SH has two transmission modes:

- An OFDM mode based on DVB-T standard with enhancements. This mode can be used on both the direct and indirect paths; the two signals are combined in the receiver to strengthen the reception in a single frequency network (SFN) configuration.
- A TDM mode partly derived from DVB-S2 standard, in order to optimize transmission through satellite towards mobile terminals. This mode is used on the direct path only. The system supports code diversity recombination between satellite TDM and terrestrial OFDM modes so as to increase the robustness of the transmission in relevant areas (mainly suburban).

The DVB-SH standard specifies the digital signal format and the digital signal modulation and coding in order to allow compatibility between pieces of equipment developed by different manufacturers. Signal processing at the modulator side is described in details, while processing at receiver side is left open to a particular implementation while complying with the standard.

1.17.1 Transmission System Architecture

The DVB-SH is mainly designed to transport mobile TV services. It may also support a wide range of mobile multimedia services, for example, audio and data broadcast as well as file download services. It performs the adaptation and transmission of one or two (in case of hierarchical mode) baseband signals to both satellite and terrestrial channel characteristics. Like DVB-S/S2, it takes MPEG transport streams (MPEG-TS) as baseband signals at system input by default.

Figure 1.33 describes the transmission system. It includes two modulation possibilities for the satellite path: an OFDM mode based on the DVB-T standard and a TDM mode, partly derived from a DVB-S2 structure.

The functional block shows a common part to both modes, and parts dedicated to each mode:

- Common functions for both modes:
 - Mode adaptation: CRC-16 and insertion of the encapsulation frame header.
 - Stream adaptation: padding and scrambling of the encapsulation frame.
 - Forward error correction (FEC) encoding using 3GPP2 turbo code.
 - Bit-wise interleaving applying on a FEC block. The latter is meanwhile shortened to comply with the modulation frame structure of OFDM and TDM.
 - Convolutional time interleaving and framing.

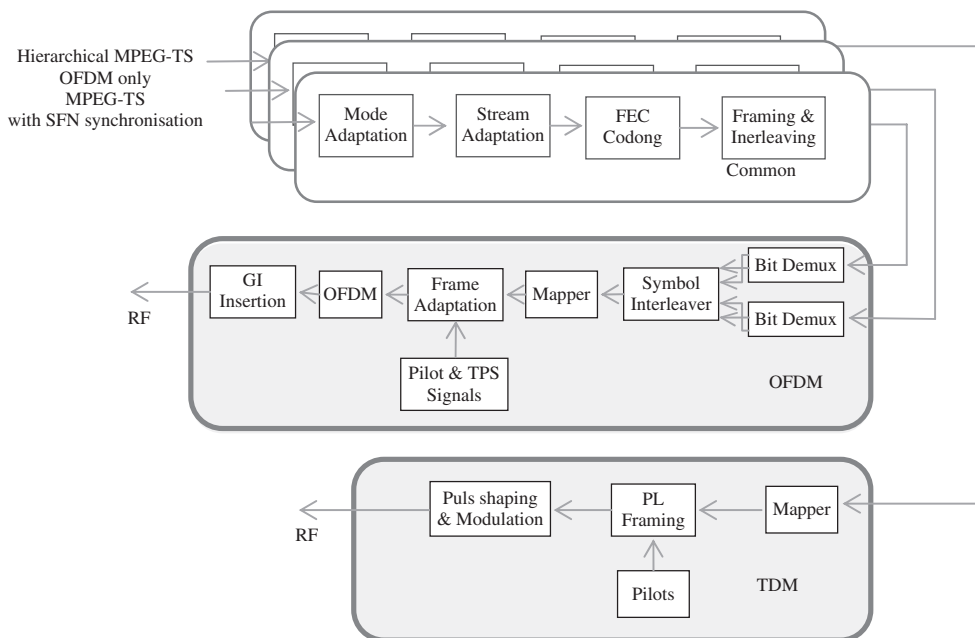


Figure 1.33 Functional block diagram of the DVB-SH transmitter (Either TDM or OFDM configurations) (Source: ETSI 2010 [13]. Reproduced with permission of ETSI.)

- Functions for the TDM mode:
 - Bit mapping to the constellation.
 - TDM physical layer framing.
 - Pilots insertion and scrambling.
 - Pulse shaping and quadrature modulation.
- Functions for the OFDM mode:
 - Symbol interleaver.
 - Bit mapping to the constellation.
 - OFDM framing with pilots and TPS insertion.

1.17.2 Common Functions for both TDM and OFDM Modes

1.17.2.1 Mode Adaptation

Mode adaptation function consists of CRC encoding (to provide error detection on every MPEG packet) and inserting an encapsulation signalling (ESignalling). It is designed to support MPEG-TS input stream, but can also handle any input stream format, be it packetized or not. The output of mode adaptation is composed of an EHEADER followed by a DATAFIELD.

1.17.2.2 Stream Adaptation

Stream adaptation provides padding to complete a constant length ($L_{\text{TC-input}} = 12\,282$ bits) encapsulation frame (EFRAME) and performs scrambling. EFRAME is designed so as to match the input turbo code block size, namely $L_{\text{TC-input}} = 12\,282$ bits, independently of the code rate.

1.17.2.3 FEC Encoding and Channel Interleaving

The DVB-S uses the turbo code for FEC as standardized by the 3GPP2 organization. Additional code rates with respect to the originally defined 3GPP2 code rates have been introduced to both allow finer granularity in terms of C/N adjustment and code combining between OFDM and TDM.

The turbo encoder employs two systematic and recursive convolutional encoders connected in parallel, with an interleaver, the turbo interleaver, preceding the second recursive convolutional encoder. During encoding, an encoder output tail sequence is added.

For any code rate, if the total number of bits encoded by the turbo encoder is $L_{\text{TC-input}}$, the turbo encoder generates $(L_{\text{TC-input}} + 6)/\text{CR}$ encoded output symbols, where CR is the code rate. The two recursive convolutional codes are called the constituent codes of the turbo code. The outputs of the constituent encoders are punctured and repeated to achieve the $(L_{\text{TC-input}} + 6)/\text{CR}$ output symbols. The $L_{\text{TC-input}}$ is set to 12 282 bits for content issued from the Stream Adaptation, and to 1146 bits for the signalling content.

1.17.2.4 Channel Interleaver and Rate Adaptation

Interleavers are used to enhance the resistance of the waveform to short-term fading and medium-term shadowing/blockage impairments in terrestrial and satellite channels. The interleaver diversity is largely provided by a common channel time interleaver.

The channel time interleaver is composed of two cascaded elementary interleavers, a block bit-wise interleaver working on individual coded words at the output of the encoder, and a convolutional time interleaver working on interleaving units (IUs) of 126 bits. A rate adaptation is inserted at the output of the bitwise interleaver in order to match the coded words to an integer number of IUs.

The bit and time interleaving processes do not depend on modulation scheme, since they are working on interleaving units. However the resulting duration of the interleaving depends on the modulation.

1.17.2.5 Frame Structure

Turbo code word framing is fully synchronized with SH frame (start of a SH frame is start of an encoded word). The bitwise interleaver followed by the rate adaptation produce Interleaving Units (IUs) of 126 bits which are fed into the time interleaver but those IU are coming from:

- the DATA only part for TDM mode;
- the DATA and PADDING parts for OFDM mode.

Those bit streams are assembled to produce SH frames.

1.17.3 Functions for Single Carrier (TDM) Mode

Combined operation of single carrier (coming from the satellite) and multi carrier (coming from a terrestrial network) has an impact on the frame parameters: to simplify the diversity reception of both signals in hybrid TDM/OFDM environment, the framing duration for the TDM waveform is made identical to the framing duration for the OFDM waveform.

Since each may use different bandwidth and FEC coding rates, this leads to different symbol and bit rates and hence capacity units. The interface to the time interleaver is the SH frame composed of the number of capacity units listed in table Table 1.7 as a function of the TDM and OFDM physical layer parameters, assuming that TDM and OFDM have the same channel bandwidth (only the 5 MHz case is presented in Table 1.7).

The signalling information is transmitted once each SH frame period. No additional TDM signalling is introduced. At the beginning of each SH frame, three capacity units carry all relevant signalling information. The signalling field is mapped like the payload data of the SH frame.

1.17.3.1 Bit Mapping into Constellation

The DVB-SH uses the QPSK, 8PSK and 16APSK constellations and the associated mapping as defined by DVB-S2 standard.

As OFDM definition relies on the DVB-T standard, TDM Frame time duration is constrained by the OFDM frame duration which value varies with bandwidth, guard interval setting and modulation order. The specified TDM framing is defined such as to cope with these frame time duration variation. The symbol rates for TDM are defined in Table 1.8. The selection of the TDM symbol rates takes into account the OFDM parameters.

Table 1.7 TDM SH FRAME transport capability in capacity units (a block 2016 bit)

OFDM guard interval	DTM roll-off	OFDM: QPSK			OFDM: 16QAM		
		TDM: QPSK	TDM: 8PSK	TDM: 16APSK	TDM: QPSK	TDM: 8PSK	TDM: 16PSK
1/4	0.15	952	1428	1904	476	714	952
1/4	0.25	896	1344	1792	448	672	896
1/4	0.35	812	1218	1624	406	609	812
1/8	0.15	868	1302	1736	434	651	868
1/8	0.25	784	1176	1568	392	588	784
1/8	0.35	728	1092	1456	364	546	728
1/16	0.15	812	1218	1624	409	609	812
1/16	0.25	756	1134	1512	378	567	756
1/16	0.35	700	1050	1400	350	525	700
1/32	0.15	784	1176	1568	392	588	784
1/32	0.25	728	1092	1456	364	546	728
1/32	0.35	672	1008	1344	336	504	672

Source: ETSI 2010 [13]. Reproduced with permission of ETSI.

1.17.3.2 TDM Framing

The SH frame to be transmitted in TDM mode consists of a number of physical layer slots (PL SLOTS) of length $L_{TOT} = 2176$ symbols, each of them comprising of 2, 3 or 4 capacity units (CU) of 2 016 bits. The capacity units are directly mapped on PL SLOTS, dependent on the modulation format as described in Table 1.9.

The number of capacity units per SH frame depends on the selection of OFDM modulation, guard interval and roll-off selection.

1.17.3.3 Pilot Insertion

There are two PILOT FIELDS (PF) of equal duration $L_{PF} = 80$ symbols in each physical layer (PL) SLOT. Each pilot symbol is an un-modulated symbol, identified by:

$$I = \frac{1}{\sqrt{2}} \text{ and } Q = \frac{1}{\sqrt{2}}$$

A PF is inserted before each SUB-SLOT (SS) of length $L_{SS} = 1\,008$ symbols. Figure 1.34 illustrates the pilot organization in the PL SLOT.

1.17.3.4 Physical Layer Scrambling

Prior to modulation, each PL SLOT including the PILOT FIELDS is randomized for energy dispersal by multiplying for each PL SLOT of length L_{TOT} the I and Q modulated baseband signal symbol samples by a unique complex randomization sequence:

$$C(i) = C_I(i) + C_Q(i), \quad I = 1, 2, \dots, L_{TOT}$$

Table 1.8 TDM symbol rates for all channelizations and as a function of the OFDM parameter settings (sampling frequency and guard interval) and of the TDM roll-off factor

Signal bandwidth in MHz	OFDM sampling frequency in MHz	OFDM guard interval	TDM symbol rate in MHz	TDM roll-off factor	TDM symbol rate in MHz	TDM roll-off factor	TDM symbol rate in MHz	TDM roll-off factor
0.8	64/7	1/4	34/5	0.15	32/5	0.25	29/5	0.35
		1/8	62/9		56/9		52/9	
		1/16	116/17		108/17		100/17	
		1/32	224/33		208/33		64/11	
0.7	8/1	1/4	119/20		28/5		203/40	
		1/8	217/36		49/9		91/18	
		1/16	203/34		189/34		175/34	
		1/32	196/33		182/33		56/11	
0.6	48/7	1/4	51/10		24/5		87/20	
		1/8	31/6		14/3		13/3	
		1/16	87/17		81/17		75/17	
		1/32	56/11		52/11		48/11	
0.5	40/7	1/4	17/4		4/1		29/8	
		1/8	155/36		35/9		65/18	
		1/16	145/34		135/34		125/34	
		1/32	140/33		130/33		40/11	
1.70	64/35	1/4	34/25		32/25		29/25	
		1/8	62/45		56/45		52/45	
		1/16	116/85		108/85		20/17	
		1/32	224/165		208/165		64/55	

Source: ETSI 2010 [13]. Reproduced with permission of ETSI.

Table 1.9 TDM framing, number of CU per PL slot

Modulation	CU per PL Slot
QPSK	2
8PSK	3
16APSK	4

Source: ETSI 2010 [13]. Reproduced with permission of ETSI.

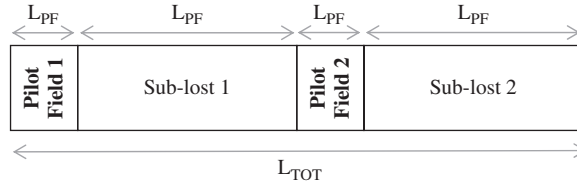


Figure 1.34 Slot pilot insertion (Source: ETSI 2010 [13]. Reproduced with permission of ETSI.)

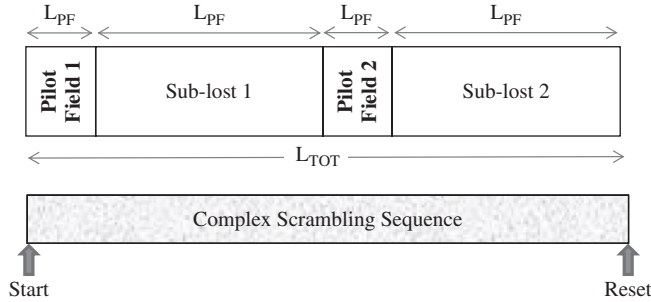


Figure 1.35 Physical layer (PL) scrambling (Source: ETSI 2010 [13]. Reproduced with permission of ETSI.)

So that:

$$I_{SCR}C(i) = I(i)C_I(i) - Q(i)C_Q(i), \quad I_{SCR}C(i) = I(i)C_I(i) + Q(i)C_Q(i), \quad I = 1, 2, \dots, L_{TOT}$$

The randomization sequence shall be reinitialized at the beginning of each PL SLOT, that is, terminated after L_{TOT} symbols as illustrated in Figure 1.35.

The scrambling code sequence is constructed by combining two real m-sequences (generated by means of two generator polynomials of degree 18) into a complex sequence. The resulting sequences thus constitute segments of a set of Gold sequences. Only one complex sequence $C_I(i) + jC_Q(i)$ is required for DVB-SH.

1.17.3.5 Baseband Shaping and Quadrature Modulation

The signals when square root raised cosine filtered with the roll-off factor shall be $\alpha = 0.15$, 0.25 and 0.35, referring to the DVB-S section for their expression.

Quadrature modulation shall be performed by multiplying the in-phase and quadrature samples (after baseband filtering) by $\sin(2\pi f_0 t)$ and $\cos(2\pi f_0 t)$, respectively (where f_0 is the carrier frequency). The two resulting signals are added to obtain the modulator output signal.

1.17.4 Functions for Multi Carrier (OFDM) Mode

The DVB-SH makes use of the multi carrier, based on the DVB-T physical layer. There are three fast Fourier transfer (FFT) modes defined in the DVB-T: 2k, 4k and 8k.

To cope with reduced signal bandwidth at L-band (channelization of 1.74 MHz), an additional 1k mode is defined. It is a strict down scaling of the existing DVB-T modes.

1.17.4.1 Interface to SH Frame

The capacity units (CU) are aligned to the OFDM symbols. An integer number of CU maps to another integer number of OFDM symbols, dependent on FFT sizes and selected subcarrier modulation. This eases the demapping of the CU and the synchronisation of the deinterleaver in the receiver. In any case, the SH frame of 816 CU is always fully aligned with the OFDM frame (see Figure 1.36).

1.17.4.2 Bit Demultiplexing

The channel interleaver produces up to two bit streams (in case of hierarchical modulation). The bit stream is demultiplexed into two sub-streams for QPSK or four for 16-QAM. In non-hierarchical mode, the single input stream is demultiplexed into ν sub-stream, where $\nu = 2$ for QPSK and $\nu = 4$ for QAM. In hierarchical mode the high priority stream is demultiplexed into two sub-streams and the low priority stream is demultiplexed into $\nu-2$ sub-streams.

1.17.4.3 Symbol Interleaver

The symbol interleaver maps ν bit words onto the 756 (1K mode), 1512 (2K mode), 3024 (4K mode) or 6048 (8K mode) active carriers per OFDM symbol. The symbol interleaver acts on vectors Y' of 756 (1K mode), 1512 (2K mode), 3024 (4K mode) or 6048 (8K mode) data symbols.

Thus in the 1K mode, a vector $Y' = (y'_0, y'_1, y'_2, \dots y'_{755})$ is assembled from 36 groups of 21 data sub words. In the 2K mode, the 72 groups of 21 words of Y' form a vector $Y' = (y'_0, y'_1, y'_2, \dots y'_{1511})$. In the 4K mode, a vector $Y' = (y'_0, y'_1, y'_2, \dots y'_{3023})$ is assembled from 144 groups of 21 data sub words. Similarly in the 8K mode, a vector $Y' = (y'_0, y'_1, y'_2, \dots y'_{6047})$ is assembled from 288 groups of 21 data sub words.

1.17.4.4 Bit Mapping into Constellation

The DVB-SH uses orthogonal frequency division multiplex (OFDM) transmission. All data carriers in one OFDM frame are modulated using QPSK, 16-QAM or non-uniform 16-QAM

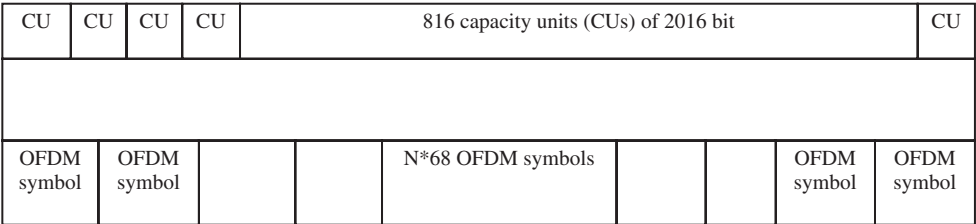


Figure 1.36 Mapping of the SH frame on OFDM (Source: ETSI 2010 [13]. Reproduced with permission of ETSI.)

constellations. The exact proportions of the constellations depend on a parameter α , which can take the three values 1, 2 or 4. α is the minimum distance separating two constellation points carrying different HP-bit values divided by the minimum distance separating any two constellation points. Non-hierarchical transmission uses the same uniform constellation as the case with $\alpha = 1$. The exact values of the constellation points are $z \in \{n + jm\}$ with values of n, m given below for the various constellations:

- **QPSK**: $n \in \{-1, 1\}, m \in \{-1, 1\}$
- **16-QAM** (non-hierarchical and hierarchical with $\alpha = 1$): $n \in \{-3, -1, 1, 3\}, m \in \{-3, -1, 1, 3\}$
- **Non-uniform 16-QAM** (with $\alpha = 2$): $n \in \{-4, -2, 2, 4\}, m \in \{-4, -2, 2, 4\}$
- **Non-uniform 16-QAM** (with $\alpha = 4$): $n \in \{-6, -4, 4, 6\}, m \in \{-6, -4, 4, 6\}$

1.17.4.5 OFDM Frame Structure

The transmitted signal is organized in frames. Each frame has duration of T_F , and consists of 68 OFDM symbols. Four frames constitute one super-frame. Each symbol is constituted by a set of $K = 6817$ carriers in the 8k mode, of $K = 3409$ carriers in the 4k mode, of $K = 1705$ carriers in the 2k mode and $K = 853$ carriers in the 1k mode, and transmitted with a duration T_S . It is composed of two parts: a useful part with duration T_U and a guard interval with duration Δ . The guard interval consists in a cyclic continuation of the useful part, T_U , and is inserted before it. Four values of guard intervals may be used.

The symbols in an OFDM frame are numbered from 0 to 67. All symbols contain data and reference information. Since the OFDM signal comprises many separately modulated carriers, each symbol can in turn be considered to be divided into cells, each corresponding to the modulation carried on one carrier during one symbol.

In addition to the transmitted data an OFDM frame contains:

- scattered pilot cells;
- continual pilot carriers;
- transmission parameter signalling (TPS) carriers.

1.17.4.6 Pilot Signals and Transmission Parameter Signalling (TPS)

The pilot signals are used for frame synchronization, frequency synchronization, time synchronization, channel estimation, transmission mode identification and can also be used to follow the phase noise.

The DVB-SH reuses the transmission parameter signalling (TPS) structure in the DVB-T standard. Compatibility is ensured via non-modification of most important parameters including modulation, hierarchy information, guard interval and transmission mode, frame number in a super-frame and cell identification. Signalling of DVB-T including options covering 4k, in-depth inner interleaver, time-slicing and MPE-FEC bits for DVB-H signalling are then used.

The TPS is transmitted in parallel on 7 TPS carriers in 1k mode, 17 carriers for the 2k mode, on 34 carriers for the 4k mode and on 68 carriers for the 8k mode. Every TPS carrier in the same symbol conveys the same differentially encoded information bit. Table 1.10 shows the carrier indices for TPS carriers.

Table 1.10 Carrier indices for TPS carriers

2k mode					8k mode							
34	50	209	346	413	34	50	209	346	413	569	595	688
569	595	688	790	901	790	901	1073	1219	1262	1286	1469	1594
1073	1219	1262	1286	1469	1687	1738	1754	1913	2050	2117	2273	2299
1594	1687				2392	2494	2605	2777	2923	2966	2990	3173
					3298	3391	3442	3458	3617	3754	3821	3977
					4003	4096	4198	4309	4481	4627	4670	4694
					4877	5002	5095	5146	5162	5321	5458	5525
					5681	5707	5800	5902	6013	6185	6331	6374
					6398	6581	6706	6799				
TPS carrier indexes for 4k mode												
34	50	209	346	413	569	595	688	790	901	1073	1219	
1262	1286	1469	1594	1687	1738	1754	1913	2050	2117	2273	2299	
2392	2494	2605	2777	2932	2966	2990	3173	3298	3391			
1k mode												
34	209	346	413									
569	688	790										

Source: ETSI 2010 [13]. Reproduced with permission of ETSI.

The TPS carriers convey information on:

- modulation including the α value of the QAM constellation pattern (see note);
- hierarchy information;
- guard interval;
- transmission mode (1k, 2k, 4k, 8k);
- frame number in a super-frame;
- cell identification;
- DVB-SH mode (selector bit);
- code rates;
- time interleaver configuration;
- super frame number in a SH frame.

Note: The α value defines the modulation based on the cloud spacing of a generalised QAM constellation. It allows specification of uniform and non-uniform modulation schemes, covering QPSK and 16-QAM.

1.17.4.7 Baseband Shaping and Quadrature Modulation

Spectrum characteristics is the same as the DVB-T. The OFDM symbols constitute a juxtaposition of equally-spaced orthogonal carriers.

The power spectral density frequency f_k at $P_k(f)$ is defined in the following expression:

$$P_k(f) = \left[\frac{\sin \pi (f - f_k) T_s}{\pi (f - f_k) T_s} \right]^2$$

where $f_k = f_c + \frac{k}{T_u}$; $\left(-\frac{K-1}{2} \leq k \leq \frac{K-1}{2}\right)$.

The overall power spectral density of the modulated data cell carriers is the sum of the power spectral densities of all these carriers. Because the OFDM symbol duration is larger than the inverse of the carrier spacing, the main lobe of the power spectral density of each carrier is narrower than twice the carrier spacing.

1.17.5 DVB-RCS2

In March 2011, two DVB-RCS2 specifications were issued for the second generation DVB interactive satellite system. DVB-RCS2 makes use of state-of-the art technology to achieve *substantial enhancements over first generation systems*. It was developed for interactive satellite services.

Details can be found in the ETSI EN 301 545-1 (V1.1.1): *Digital video broadcasting; second generation DVB interactive satellite system (DVB-RCS2)*, part 1: *overview and system level specification*, and part 2: *lower layer for satellite standard*, 05/2012.

DVB-S2 is based on DVB-S systems for internetworking between the satellite with small or large networks as well as fixed or mobile terminals. It takes advantages of the evolutions of the physical layer techniques and the stabilisation of IP standards for consistent implementation.

The first part of the multi-part specification of that system provides an overview of the system has been completed. It contains the requirements in order to provide the best possible interoperability between terminals and hubs, thus defining not only the lower layers of the system (up to layer 2) but also network functions as well as management and control capabilities.

As these specifications combined together may end up in a very complex terminal design, the specification also describes subsets of capabilities known as profiles that can be used together to address a given market segment.

The second part specified the lower layers and the lower layer embedded signalling for the management and control system, for two-way interactive satellite networks.

It allows substantial configuration flexibility in that the burst constructions and FEC can be adapted to some extent to the operating environment of the satellite terminal.

1.18 Historical Development of Computer and Data Networks

Telecommunication systems and broadcasting systems have been developing for over 100 years. The basic principles and services have changed little since their beginnings and we can still recognise the earliest telephony systems and televisions. However, computers and the Internet have changed greatly in the last 40 years. Today's systems and terminals are completely different from those used 40 or even 10 years ago. The following gives a quick review of these developments to show the pace of technology progress.

1.18.1 Dawn of the Computer and Data Communications Age

The first electronic digital computer was developed during 1943–1946. Early computer interfaces used punched tapes and cards. Later terminals were developed and the first communication between terminals and computer over long distances was in 1950, which used voice-grade telephone links at low transmission speeds of 300–1200 kbit/s. Automatic repeat requests (ARQ) for error correction were mainly used for data transmission.

1.18.2 Development of Local Area Networks (LANs)

From 1950 to 1970 research carried out on computer networks led to the development of different types of network technologies – local area networks (LANs), metropolitan area networks (MANs) and wide area networks (WANs).

A collection of standards, known as IEEE 802, was developed in the 1980s including the Ethernet as IEEE802.3, token bus as IEEE802.4, token ring as IEEE802.5, DQDB as IEEE802.6 and others. The initial aim was focused on LAN and MAN to share file systems and expensive peripheral devices such as high-quality printers and graphical plot machines at fast data rates. The WAN was left to the ISO and ITU-T to develop the B-ISDN for interconnection of the LANs and MANs.

1.18.3 Development of WANs and ISO/OSI

The ISO developed the open system interconnection (OSI) reference model with seven layers for use in wide area networks in the 1980s. The goal of the reference model was to provide an open standard so that different terminals and computer systems could be connected together if they conformed to the standard. The terminals considered in the reference model were connected to a mainframe computer over a WAN in text mode and at slow speed.

1.18.4 Birth of the Internet

Many different network technologies were developed during the 1970s and 1980s and many of them did not fully conform to international standards. Internetworking between different types of networks used protocol translators and interworking units, and became more and more complicated as the protocol translators and interworking units became more technology dependent.

In the 1970s, the Advanced Research Project Agency Network (ARPANET) sponsored by the United States Department of Defense developed a new protocol, which was independent of network technologies, to interconnect different types of networks. The ARPANET was renamed as the Internet in 1985. The main application layer protocols included remote telnet for terminal access, FTP for file transfer and email for sending mail through computer networks.

1.18.5 Integration of Telephony and Data Networks

In the 1970s, the ITU-T started to develop standards called integrated services digital networks with end-to-end digital connectivity to support a wide range of services, including voice and

non-voice services. User access to the ISDN was through a limited set of standard multipurpose customer interfaces. Before ISDN, access networks, also called local loops, to the telecommunication networks were analogue, although the trunk networks, also called transit networks, were digital. This was the first attempt to integrate telephony and data networks and integration of services over a single type of network. It still followed the fundamental concepts of channel- and circuit-based networks used in traditional telecommunication networks.

1.18.6 Development of Broadband Integrated Networks

As soon as the ISDN was completed in the 1980s, the ITU-T started to develop broadband ISDN. In addition to broadband integrated services, ATM technology was developed to support the services based on fast packet-switching technologies. New concepts of virtual channels and virtual circuits were developed. The network is connection oriented, which allows negotiation of bandwidth resources to meet the QoS requirements of different types of services and applications. It was expected to unify the telephony networks and data networks and also unify LANs, MANs and WANs.

From the LAN aspect, ATM faced fierce competition from fast and Gigabit Ethernet. From application aspects, it faced competition from the Internet.

1.18.7 The Killer Application WWW and Internet Evolutions

In 1990, Tim Berners-Lee developed a new application called the World Wide Web (WWW) based on hypertext over the Internet. This significantly changed the direction of network research and development. A large number of issues needed to be addressed to cope with the requirements of new services and applications, including real-time services and their quality of service (QoS), which were not considered in traditional Internet applications.

In addition, all services and applications are evolving towards all IP solutions such as computer, telephone, mobile phone and television. New protocols were needed to accommodate these services by introducing new features such as QoS, security, multicast, broadcast, large address space and evolution strategies. The next generation IP, called IPv6 was developed and standardised to off these new features. Officially, the deployment and evolution toward IPv6 started on 6 June 2012.

1.19 Historical Development of Satellite Communications

Satellite has been associated with telecommunications and television from its beginning, through few people have noticed this. Today, satellites broadcast television programmes directly to our homes and allow us to transmit messages and surf the Internet. The following gives a quick review of satellite history.

1.19.1 Start of Satellite and Space Eras

Satellite technology has advanced significantly since the launch of the first artificial satellite *Sputnik* by the USSR on 4 October 1957 and the first experiment of an active relaying communications satellite *Courier-1B* by the USA in August 1960.

The first international cooperation to explore satellite for television and multiplexed telephony services was marked by the experimental pre-operation transatlantic communications between the USA, France, Germany and the UK in 1962.

1.19.2 Early Satellite Communications: TV and Telephony

Establishment of the Intelsat – International Satellite Organisation started with 19 national administration and initial signatories in August 1964. The launch of the REARLY BIRD (Intelsat-1) marked the first commercial geostationary communication satellite. It provided 240 telephone circuits and one TV channel between the USA, France, Germany and the UK in April 1965. In 1967, Intelsat-II satellites provided the same service over the Atlantic and Pacific Ocean regions. From 1968 to 1970, Intelsat-III achieved worldwide operation with 1500 telephone circuits and four TV channels. The first Intelsat-IV satellite provided 4000 telephone circuits and two TV channels in January 1971 and Intelsat-IVa provided 20 transponders of 6000 circuits and two TV channels, which used beam separation for frequency reuse.

1.19.3 Development of Satellite Digital Transmission

In 1981, the first Intelsat-V satellite achieved capacity of 12 000 circuits with FDMA and TDMA operations, 6/4 GHz and 14/11 GHz wideband transponders, and frequency reuse by beam separation and dual polarisation. In 1989, the Intelsat-VI satellite provided onboard satellite-switched TDMA of up to 120 000 circuits. In 1998, Intelsat VII, VIIa and Intelsat-VIII satellites were launched. In 2000, the Intelsat-IX satellite achieved 160 000 circuits. On 18 July 2001, Intelsat became a private company. It was also the time when satellite networks were evolving toward packet-oriented broadband and TV distribution services.

1.19.4 Development of Direct-to-Home (DTH) Broadcast

In 1999, the first Ku band TV satellite provided 30 14/11–12 GHz transponders for 210 TV programmes with possible direct-to-home (DTH) broadcast and VSAT services. Due to VSAT with small antenna of 0.75–1.2 m, it is possible to provide a large number of user terminals at 56 kbit/s initially, and up to a few Mbit/s today. It also evolves towards broadband access services via satellites.

1.19.5 Development of Satellite Maritime Communications

In June 1979, the International Maritime Satellite (Inmarsat) organisation was established to provide global maritime satellite communication with 26 initial signatories. It explored the mobility feature of satellite communications. It was private in 1999. It also provides services to aircraft and portable terminals for land mobile users. Through its broadband global area networks (BGAN), it provides satellite broadband integrated services to remote users for laptop the Internet access as well as telephony services.

1.19.6 Satellite Communications in Regions and Countries

At a regional level, the European Telecommunication Satellite (Eutelsat) organisation was established with 17 administrations as initial signatories in June 1977. Many countries also

developed their own domestic satellite communications systems, including the USA, the USSR, Canada, France, Germany, the UK, Japan, China and other nations. Eutelsat provided its services when the first satellite launched in 1983. With the waves of privatization in the telecommunication sector, Eutelsat was privatised in 2001, becoming a private company called Eutelsat S.A.

1.19.7 Satellite Broadband Networks and Mobile Networks

Since the 1990s, significant development had been carried out on broadband networks including onboard-switching satellite technologies. Various non-geostationary satellites have been developed for mobile satellite services (MSSs) and broadband fixed satellite services (FSSs). MSS requires more powerful satellite for GEO or a constellation of LEO satellites for global continuously coverage.

1.19.8 Internet over Satellite Networks

Since the late 1990s and the start of the twenty-first century, we have seen a dramatic increase in Internet traffic over the communication networks. Satellite networks have been used to transport Internet traffic in addition to telephony and television traffic for access and transit networks. This brings great opportunities as well as challenges to the satellite industry. On one hand, it needs to develop internetworking with many different types of legacy networks; and on the other hand, it needs to develop new technologies to internetwork with future networks. We have also see the convergence of different types of networks including network technologies, network protocols and new services and applications.

To explore Ka band, the SPACEWAY-3 was launched on 14 august 2007 in the United States and Ka-Sat in Europe launched on 26 December 2010. Many new satellites have been scheduled to launch in the new future due to the large bandwidth capacity and small user terminals.

1.20 Convergence of Network Technologies and Protocols

The convergence is the natural progression of technologies pushing and user demands pulling and the development of business cases. Obviously, satellite networking closely follows the development of terrestrial networks, but is capable of overcoming geographical barriers and the difficulty of wide coverage faced by terrestrial networks. Figure 1.37 illustrates the vision of a future satellite network in the context of the global information infrastructure.

1.20.1 Convergence of Services and Applications in User Terminals

In the early days, user terminals were designed for particular types of services and had very limited functions. For example, we had telephone handsets for voice services, computer terminals for data services, and television for receiving television services. Different networks were developed to support these different types of terminals.

As the technology developed, additional terminals and services were introduced into the existing networks. For example, fax and computer dialup services were added to

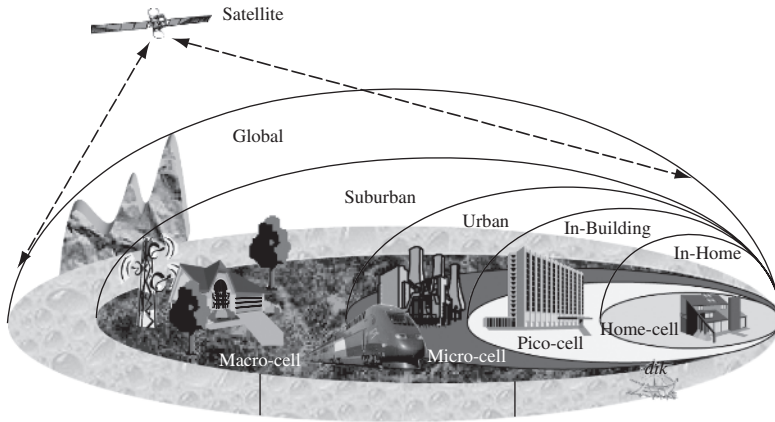


Figure 1.37 Satellite in the global information infrastructure

telephone networks. However, the transmission speeds of fax and dialup links were limited by the capacity of the telephone channel supported by the telephony networks, hence leading to development of broadband networks.

Computer terminals have become more and more sophisticated and are now capable of dealing with voice and video services in real time. Naturally, in addition to data services, there are increasing demands to support real time voice and video over data networks.

Multimedia services, a combination of voice, video and data, were developed. These complicate the QoS requirements requiring complicated user terminal and network design, implementation and operation.

To support such services over satellite networks for applications such as aeronautics, shipping, transport, emergency and disaster rescue services brings even more challenges. We are starting to see the convergence of different user terminals for different types of services into a single user terminal for all types of services. The new generation of mobile phones provides both functions of telephone and personal computers.

1.20.2 Convergence of Network Technologies

Obviously, network services are closely related to the physical networks. To support a new generation of services we need a new generation of networks. However, the design of new services and networks needs significant amounts of investment and a long period of time for research and development. To get users to accept new services and applications is also a great challenge.

How about building new services on the existing networking infrastructure? Yes, this approach has been tried as far as possible, as mentioned previously, fax and computer dialup were added to telephony networks, and voice and video services to data networks. This approach does not ease the task of developing new services and networks, as the original designs of the networks were optimised for original services. Therefore, new networks have to be developed for new services and applications.

Luckily we do not need to start from scratch. The telephone networks and services on the existing networks have been developing over the past 100 years; during that time we have accumulated a huge amount of knowledge and experience.

High and high speed of networks have been developed to meet the increasing demand of the existing as well as new services and applications.

1.20.3 Convergence of Network Protocols

Following the concepts of telecommunication networking principles, attempts were made to develop new services and networks. Examples of these were the integrated services digital networks (ISDN), synchronous transfer mode (STM) networks, broadband ISDN (B-ISDN) and asynchronous transfer mode (ATM) networks. As telephony services were historically the major services in the telecommunication networks, the new networks were biased towards these services, and were perhaps emphasising too much on real time and QoS. The results were not completely satisfactory.

Computer and data networks have been developing for about 50 years, during this time we have also accumulated a significant amount of knowledge and experience in the design of computers and data networks. All the computer and data network technologies have converged to the Internet technologies. In LAN, Ethernet is the dominating technology; other LANs, such as token ring and token bus networks are disappearing. Of course, wireless LANs are becoming popular. The Internet protocols are now the protocol for computer and data networking supporting the development of the Internet services and applications. One of the most important successful factors is backward compatibility, that is, new network technology should be capable of supporting the existing services and applications and internetworking with the existing user terminals and network without any modifications.

Following the success of the Internet, significant research and development have been carried out to support telephony services and other real-time services. As the original design of the Internet was for data services, it was optimised for reliable data services without much thought given to real-time services and QoS. Therefore, IP telephony can be achieved with the level of QoS provided by a new generation of the networks, as well as Internet TV.

Convergence of network design is inevitable, however, we have to learn from the telecommunication networks for QoS and reliable transmission of data from the Internet.

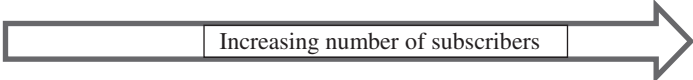
The principles of networking are still the same: to improve reliability; increase capacity; support integrated services and applications; to reach anywhere and anytime; and particularly important for satellite networking to fully utilise limited resources and reduce costs.

1.20.4 Satellite Network Evolution

It can be seen that satellite communication started from telephony and TV broadcast terrestrial networks. It went on to increase capacity, extend coverage to the oceans and air by reducing the terminal size for mobile phones and extend services to data and multimedia services.

Satellites have become more sophisticated, and have progressed from single transparent satellites to onboard processing and onboard switching satellites, and further to non-geostationary satellite constellations with inter-satellite links (ISL).

Table 1.11 Evolution of satellite broadband systems

Characterisation	First generation	Second generation	Third generation
Timeframe of operation	1980 to mid-2005	Mid-2005 to 2010	2010 to present
Satellite capacity	1 Gbps	10 Gbps	100 Gbps
			
Typical data rate per terminal	56–256 kbps	256 kbps – 3 Mbps	2–3 Mbps
			
Maximum number of subscribers per satellite	100 000–500 000	750 000–1 000 000	2 million to 3 million
			
Satellite	All FSS satellites; for example: Hughes leases 126 xpntrs worldwide	IPStar; SES Astra2Connect; Eutelsat Tooway, Wildblue, Telesat; Spaceway	ViaSat; Ka-Sat; KaComm; Spaceway 4
Satellite payload characteristics	24 Ku-band transponder with regional coverage and 36-72 MHz bandwidth	Ku-band and Ka-band spot beams with 36-72 MHz bandwidth	Ka-band spot beams with 500 MHz bandwidth
Major VSAT terminal suppliers	Hughes, Gilat, ViaSat, iDirect	Hughes, Gilat, ViaSat, iDirect	Hughes, Gilat, ViaSat, iDirect
Cost of VSAT terminal	US\$ 5000–10 000	US\$ 500–1 000	US\$ 500
Typical applications	Point of sale transaction	Broadband access for enterprise and consumer	Broadband access for enterprise and consumer
Data protocol	Proprietary and non-IP based	IP based	IP based
Connection Type	Bursty; non-real-time data	Continuous VoIP and video streaming capability	Continuous VoIP and video streaming capability

Source: Ofcom, Report No: 72/11/R/193/R

Basic satellites have a repeater to relay signals from one side to the other, called transparent pent-pipe satellites as they simply provide links between terminals without processing.

Some satellites have onboard processing (OBP) as part of the communication subsystems to provide error detection and error correction to improve the quality of the communication links, and some have onboard switching (OBS) to form a network node in the sky to explore efficient use of radio resources. Experiments have also been carried out to fly IP router onboard satellites due to the recent rapid development of Internet.

Satellites have played an important role in telecommunications networks supporting telephony, video, broadcast, data, broadband and Internet services and have become an important integrated part of the global information infrastructure providing the next generation of integrated broadband and Internet network. Table 1.11 shows the evolution of satellite broadband systems.

The future satellite systems will try to explore high frequency band to achieve data rates upto Mbit/s and even Tbit/s with much smaller transportable and mobile terminals.

Further Readings

- [1] Brady, M. and M. Rogers, Digital Video Broadcasting Return Channel via Satellite (DVB-RCS) Background Book, Nera Broadband Satellite AS (NBS), 2002.
- [2] Eutelsat, Overview of DVB, Annex B to Technical Guide, June 1999.
- [3] Haykin, S., Communication Systems, 4th edition, John Wiley & Sons, Inc., 2001.
- [4] ITU, Handbook on Satellite Communications, 3rd edition, John Wiley & Sons, Inc., 2002.
- [5] Joel, A., Retrospective: telecommunications and the IEEE communications society, IEEE Communications, May 2002.
- [6] Khader, M. and W.E. Barnes, Telecommunications Systems and Technology, Prentice-Hall, 2000.
- [7] Howard Tripp, Alan Ford, B. G. Evans, P. T. Thompson, Z. Sun and K Tara Smith, Understanding Satellite Broadband Quality of Experience – Final Report, Issue 1, Produced for: Ofcom, Report No: 72/11/R/193/R, July 2011.
- [8] ETSI EN 300 421 (V.1.1.2): Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for 11/12 GHz satellite services, 08/1997.
- [9] ETSI EN 301 210 (V.1.1.1): Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for Digital Satellite News Gathering (DSNG) and other contribution applications by satellite, 03/1999.
- [10] ETSI TR 101 154 (V.1.9.1): Digital Video Broadcasting (DVB); Implementation guidelines for the use of MPEG-2 Systems, Video and Audio in satellite, cable and terrestrial broadcasting applications, 09/1999.
- [11] ETSI EN 301 790 (V.1.5.1): Digital Video Broadcasting (DVB); Interaction channel for satellite distribution systems, 05/2009.
- [12] ETSI EN 302 307 (V.1.3.1): Digital Video Broadcasting (DVB); Second generation framing structure, channel coding and modulation system for Broadcast, Interactive Services, News Gathering and other broadband satellite applications, 03/2013.
- [13] ETSI EN 302 583 (V.1.1.2): Digital Video Broadcasting (DVB); framing structure, channel coding and modulation system for satellite services to handheld devices (SH) below 3 GHz, 02/2010.
- [14] ETSI TS 101 545-1 (V.1.1.1): Digital Video Broadcasting; Second Generation DVB Interactive Satellite System (DVB-RCS2), Part 1: Overview and System Level Specification, 05/2012.
- [15] ETSI EN 301 545-2 (V.1.1.1): Digital Video Broadcasting; Second Generation DVB Interactive Satellite System (DVB-RCS2), Part 2: Lower Layer for Satellite Standard, 05/2012.
- [16] ETSI TS 101 545-3 (V.1.1.1): Digital Video Broadcasting; Second Generation DVB Interactive Satellite System (DVB-RCS2), Part 3: Higher Layers Satellite Specification, 05/2012.
- [17] ITU-T Recommendation G.1000, Communications quality of service: A framework and definitions, 11/2001.
- [18] ITU-T Recommendation G.1010, End-user multimedia QoS categories, 11/2001.

Exercises

1. Explain the meaning of broadband, using the definition given in the ITU-T recommendations.
2. Explain the basic concepts of satellite networking and internetworking with terrestrial networks.
3. Explain the terms satellite services, network services and quality of service (QoS).
4. Discuss the differences between satellite networking and terrestrial networking issues.
5. Explain the functions of network user terminals and satellite terminals.
6. Derive the Shannon power limit and the Shannon bandwidth capacity for large E_b/N_0 .
7. Explain the basic principles of protocols and the ISO reference model.
8. Explain the basic ATM reference model.
9. Explain the Internet protocol TCP/IP suite.
10. Explain the basic concepts of multiplexing and multiple accessing.
11. Explain the basic switching concepts including circuit switching, virtual circuit switching and routing.
12. Explain the evolution process and convergence of network technologies and protocols.
13. List the recent major trends of broadband satellite systems