Chapter 2

Packet Switched Networks

2.1 Open Systems Interconnection (OSI) Model

2.1.1 History of OSI

The Open Systems Interconnection (usually abbreviated to OSI) was an effort to standardize networking that was started in 1982 by the International Organization for Standardization (ISO), along with the ITU-T.

Prior to OSI, according to its proponents, networking was largely vendor-developed and proprietary, with protocol standards such as SNA, Appletalk, NetWare and DECnet. OSI was an industry effort, attempting to get everyone to agree to common network standards to provide multi-vendor interoperability. It was common for large networks to support multiple network protocol suites, with many devices unable to talk to other devices because of a lack of common protocols between them. However while OSI developed its networking standards, TCP/IP came into widespread use on multi vendor networks, while below the network layer, both Ethernet and token ring played much the same role.

The OSI reference model (which actually predates the OSI protocol work, dating to 1977) was a major advance in the teaching of network concepts. It promoted the idea of a common model of protocol layers, defining interoperability between network devices and software.

However, the actual OSI protocol suite that was specified as part of the project was considered by many to be too complicated and to a large extent unimplementable. Taking the "forklift upgrade" approach to networking, it specified eliminating all existing protocols and replacing them with new ones at all layers of the stack. This made implementation difficult, and was resisted by many vendors and users with significant investments in other network technologies. In addition, the OSI protocols were specified by committees filled with differing and sometimes conflicting feature requests, leading to numerous optional features; because so much was optional, many vendors' implementations simply could not inter operate, negating the whole effort. Even demands by the USA for OSI support on all government purchased hardware did not save the effort.

Beyond their objections to the Byzantine protocol suite itself, OSI opponents generally contended that the very OSI standardization process represented little more than institutional unwillingness on the part of the ISO and ITU-T to admit that vendor-neutral standards might exist that had not been developed and ratified by their own particular processes. Much bad blood arose between these standards organizations and the IETF, the Internet standards body, as a result of this dispute. The most vitriolic opponents of the OSI suite at times made the further claim that, far from being a "vendor neutral" standard, OSI represented an attempt by minor, or diminishing, players in the networking and computer industries to recover by government fiat market share that they were rapidly losing to the proponents of

TCP/IP through fair competition.

The OSI approach was eventually eclipsed by the Internet's TCP/IP protocol suite. TCP/IP's pragmatic approach to computer networking and two independent implementations of simplified protocols made it a practical standard. For example, the definition for OSI's X.400 [1] e-mail standards took up several large books, while the Internet e-mail (SMTP) definition took only a few dozen pages in RFC-821. It should be noted, however, that over time there have been numerous RFCs which extended the original SMTP definition, so that its complete documentation finally took up several large books as well. Furthermore, the X.400 standard contained so many optional format choices that early implementations in France and Germany were unable to parse each other's messages.

The collapse of the OSI project in 1996 severely damaged the reputation and legitimacy of the organizations involved, especially ISO. The worst part was that OSI's backers took too long to recognize and accommodate the dominance of the TCP/IP protocol suite. The financial damage done to Japan and Europe (where Internet deployment was delayed by years) is difficult to estimate.

Philosophically, successful standards are generally defacto, based on existing proven technology. The OSI project was criticized for violating that tradition and attempting to design a complex system from scratch, by committee.

[1] X.400 is a suite of ITU-T Recommendations that define standards for Data Communication Networks for Message Handling Systems (MHS).

2.1.2 Layer 7: Application Layer

The Application layer provides a means for the user to access information on the network through an application. This layer is the main interface for the user(s) to interact with the application and therefore the network. Some examples of application layer protocols include Telnet, applications which use File Transfer Protocol (FTP), applications which use Simple Mail Transfer Protocol (SMTP) and applications which use Hypertext Transfer Protocol (HTTP). Applications built to use a protocol, such as FTP, should not be confused with the protocols themselves, which often reside at the session layer.

2.1.3 Layer 6: Presentation Layer

The Presentation layer transforms data to provide a standard interface for the Application layer. Different applications use different terminals, control characters to specify backspace, line feed, and carriage return. Also different application programs may follow different rules to encode data structures such as matrices and complex numbers. This set of rules for representing information is called *syntax*. Thus different computers use different syntaxes. The syntax used by a computer is its *local syntax*. Presentation Layer is responsible for the conversion between the local and transfer syntax. MIME encoding, data compression, data encryption and similar manipulation of the presentation is done at this layer to present the data as a service or protocol developer sees fit. Examples of this layer are converting an Extended Binary Coded Decimal Interchange Code (EBCDIC)-coded text file to an American Standard Code for Information Interchange (ASCII)-coded file, or serializing objects and other data structures into and out of Extensible Markup Language (XML).

Multipurpose Internet Mail Extensions (MIME)

MIME is an Internet Standard that extends the format of e-mail to support text in character sets other than US-ASCII, non-text attachments, multi-part message bodies, and header information in non-ASCII character sets. Virtually all human-written Internet e-mail and a fairly large proportion of automated e-mail is transmitted via SMTP in MIME format. Internet e-mail is so closely associated with the SMTP and MIME standards that it is sometimes called SMTP/MIME e-mail.

The content types defined by MIME standards are also of importance outside of e-mail, such as in communication protocols like HTTP for the World Wide Web. HTTP requires that data be transmitted in the context of e-mail-like messages, even though the data may not actually be e-mail.

MIME is specified in five RFCs : RFC 2045, RFC 2046, RFC 2047, RFC 2048 and RFC 2077.

2.1.4 Layer 5: Session Layer

The Session layer controls the dialogs/connections (sessions) between computers. It establishes, manages and terminates the connections between the local and remote application. It provides for either full-duplex or half-duplex operation, and establishes check pointing, adjournment, termination, and restart procedures. The OSI model made this layer responsible for "graceful close" of sessions, which is a property of TCP, and also for session check pointing and recovery, which is not usually used in the Internet protocols suite.

2.1.5 Layer 4: Transport Layer

The Transport layer provides transparent transfer of data between end users, thus relieving the upper layers from any concern while providing reliable data transfer. The transport layer controls the reliability of a given link through flow control, segmentation/de-segmentation, and error control. Some protocols are state and connection oriented. This means that the transport layer can keep track of the packets and retransmit those that fail. The best known example of a layer 4 protocol is the Transmission Control Protocol (TCP)[2]. The transport layer is the layer that converts messages into TCP segments or User Datagram Protocol (UDP), Stream Control Transmission Protocol (SCTP), etc. packets. Perhaps an easy way to visualize the Transport Layer is to compare it with a Post Office, which deals with the dispatching and classification of mail and parcels sent.

User Datagram Protocol (UDP)

The User Datagram Protocol (UDP) is one of the core protocols of the Internet protocol suite. Using UDP, programs on networked computers can send short messages sometimes known as datagrams (using Datagram Sockets) to one another. UDP is sometimes called the Universal Datagram Protocol or Unreliable Datagram Protocol.

UDP does not provide the reliability and ordering that TCP does. Datagrams may arrive out of order, appear duplicated, or go missing without notice. Without the overhead of checking whether every packet actually arrived, UDP is faster and more efficient for many lightweight or time-sensitive purposes. Also, its stateless nature is useful for servers that answer small queries from huge numbers of clients. Compared to TCP, UDP is required for broadcast (send to all on local network) and multicast (send to all subscribers).

[2] Refer to Chapter-3

Common network applications that use UDP include the Domain Name System (DNS), streaming media applications such as IPTV, Voice over IP (VoIP), Trivial File Transfer Protocol (TFTP) and online games.

2.1.6 Layer 3: Network Layer

The Network layer provides the functional and procedural means of transferring variable length data sequences from a source to a destination via one or more networks while maintaining the quality of service requested by the Transport layer. The Network layer performs network routing functions, and might also perform segmentation/de-segmentation, and report delivery errors. Routers operate at this layersending data throughout the extended network and making the Internet possible. This is a logical addressing scheme values are chosen by the network engineer. The addressing scheme is hierarchical. The best known example of a layer 3 protocol is the Internet Protocol (IP). Perhaps it's easier to visualize this layer as the actual Air Mail or Consolidated Carrier that transfers the mail from Point A to Point B.

Thus the network layer uses the transmission over the point-point links provided by the data link layer to transmit packets between any two computers attached in the network.

The Figure-2.1 shows a router attached to several links. When the router receives a packet, it must decide on the basis of the network addresses along which link it should retransmit the packet. This function is implemented by the network layer.

Observe that the link between, say C and D in the Figure-2.1 may carry packets between A and E and between B and E. These packets are statistically multiplexed by the router C. The network addresses of the packets permit demultiplexing.

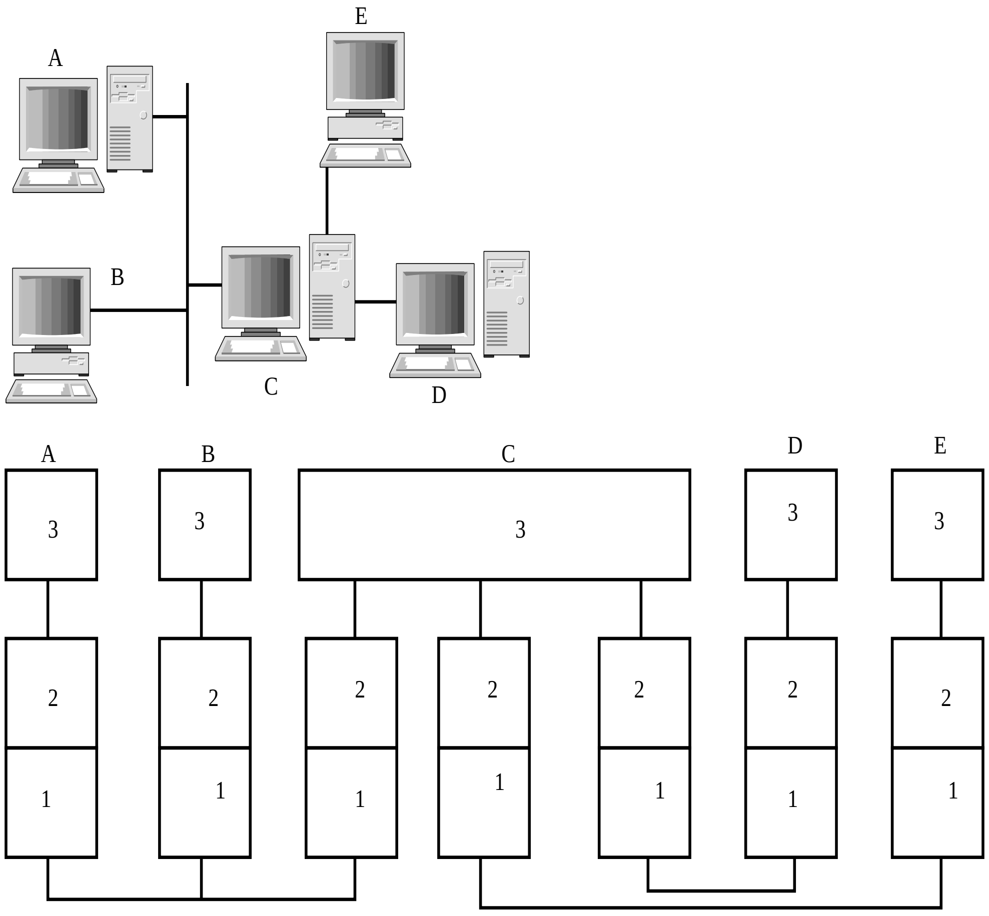


Figure 2.1: Computer C in this figure is a router. It is designed to relay packets at a high rate to the proper link and with proper delay

2.1.7 Layer 2: Data Link Layer

The Data Link layer provides the functional and procedural means to transfer data between network entities and to detect and possibly correct errors that may occur in the Physical layer. The best known example of this is Ethernet. Other examples of data link protocols are High-Level Data Link Control (HDLC) and Advanced Data Communication Control Procedures (or Protocol) (ADCCP) for point-to-point or packet-switched networks and Aloha for local area networks. On IEEE 802 local area networks, and some non-IEEE 802 networks such as FDDI, this layer may be split into a Media Access Control (MAC) layer and the IEEE 802.2 Logical Link Control (LLC) layer. It arranges bits from physical layer into logical chunks of data, known as frames.

This is the layer at which the bridges and switches operate. Connectivity is provided only among locally attached network nodes forming layer 2 domains for unicast or broadcast forwarding. Other protocols may be imposed on the data frames to create tunnels and logically separated layer 2 forwarding domain.

2.1.8 Layer 1: Physical Layer

The Physical layer defines all the electrical and physical specifications for devices. This includes the layout of pins, voltages, and cable specifications. Hubs, repeaters, network adapters and Host Bus Adapters (HBAs used in Storage Area Networks) are physical-layer devices. The major functions and services performed by the physical layer are:

- Establishment and termination of a connection to a communications medium.

- Participation in the process whereby the communication resources are effectively shared among multiple users. For example, contention resolution and flow control.

- Modulation, or conversion between the representation of digital data in user equipment and the corresponding signals transmitted over a communications channel. These are signals operating over the physical cabling (such as copper and fiber optic) or over a radio link.

Parallel Small Computer System Interface (SCSI) buses operate in this layer. Various physical-layer Ethernet standards are also in this layer; Ethernet incorporates both this layer and the data-link layer. The same applies to other local-area networks, such as Token ring, FDDI, and IEEE 802.11, as well as personal area networks such as Bluetooth and IEEE 802.15.4.

OSI Model			
	Data Unit	Layer	Function
Host Layers	Data	Application	Network process to application
		Presentation	Data representation and encryption
		Session	Interhost communication
	Segments	Transport	End-to-end connections and reliability (TCP)
Media Layers	Packets	Network	Path determination and logical addressing (IP)
	Frames	Data link	Physical addressing (MAC and LLC)
	Bits	Physical	Media, signal and binary transmission

Table 2.1: Summary of OSI Model

2.2 Ethernet (IEEE 802.3)

Ethernet has pretty much taken over the LAN market. As recently as the 1980s and the early 1990s, Ethernet faced many challenges from other LAN technologies, including token ring, FDDI and ATM. Some of these other technologies succeeded at capturing a part of the market share for a few years. But since its invention in the mid-1970, Ethernet has continued to evolve and grow, and has held on to its dominant market share. Today, Ethernet is by far the most prevalent LAN technology, and is likely to remain so for the foreseeable future. One might say that Ethernet has been to local area networking what the Internet has been to global networking:

There are many reasons for Ethernet's success. First, Ethernet was the first widely-deployed high-speed LAN. Because it was deployed early, network administrators became intimately familiar with Ethernet – its wonders and its quirks – and were reluctant to switch over to other LAN technologies when they came on the scene. Second, token ring, FDDI and ATM are more complex and expensive than Ethernet, which further discouraged network administrators from switching over. Third, the most compelling reason to switch to another LAN technology (such as FDDI or ATM) was usually the higher data rate of the new technology; however, Ethernet always fought back, producing versions that operated at equal data rates or higher. Switched Ethernet was also introduced in the early 1990s, which further increased its effective data rates. Finally, because Ethernet has been so popular, Ethernet hardware (in particular, network interface cards) has become a commodity and is remarkably cheap. This low cost is also due to the fact that Ethernet's multiple access protocol, CSMA/CD, is totally decentralized, which has also contributed to the low cost and simple design. The original Ethernet LAN, as shown in Figure-2.2 , was invented in the mid 1970s by Bob Metcalfe.

Physical Layer

The first Ethernet networks, 10BASE5, used thick yellow cable with vampire taps as a shared medium (using carrier sense multiple access with collision detection (CSMA/CD)). Later, 10BASE2 Ethernet used thinner coaxial cable (with BNC connectors) as the shared CSMA/CD medium. The later StarLAN 1BASE5 and 10BASE-T used twisted pair connected to Ethernet hubs with 8P8C modular connectors (not to be confused with FCC's RJ45).

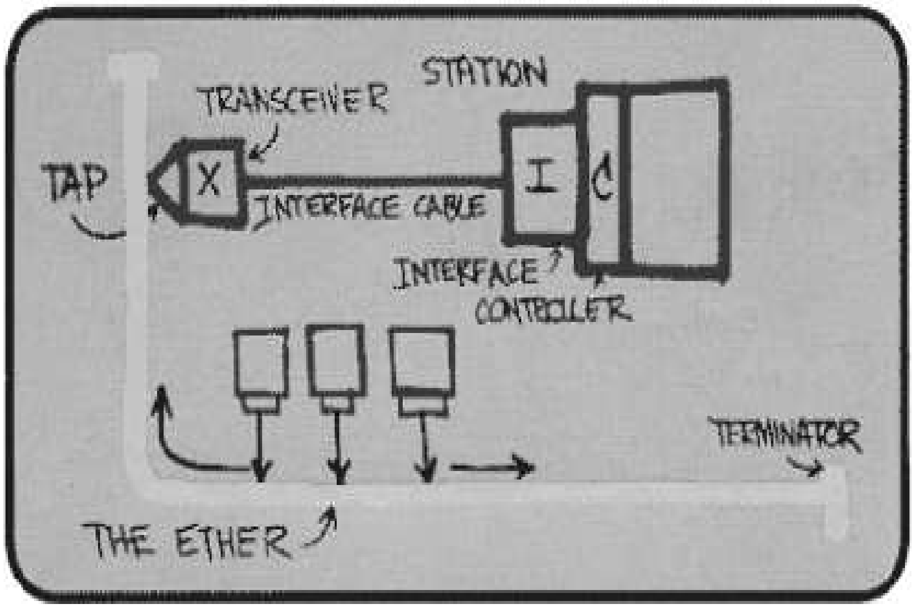


Figure 2.2: The original Metcalfe design led to the 10Base5 Ethernet standard, which included an interface cable that connected the Ethernet adapter (i.e., interface) to an external transceiver. Drawing taken from Charles Spurgeon's Ethernet Web Site.

Currently Ethernet has many varieties that vary both in speed and physical medium used. Perhaps the most common forms used are 10BASE-T, 100BASE-TX, and 1000BASE-T. All three utilize twisted pair cables and 8P8C modular connectors (often incorrectly called RJ45). They run at 10 Mbit/s, 100 Mbit/s, and 1 Gbit/s, respectively. However each version has become steadily more selective about the cable it runs on and some installers have avoided 1000BASE-T for everything except short connections to servers.

Fiber optic variants of Ethernet are commonly seen connecting buildings or network cabinets in different parts of a building but are rarely seen connected to end systems for cost reasons. Their advantages lie in performance, electrical isolation and distance, up to tens of kilometers with some versions. Fiber versions of a new speed almost invariably come out before copper. 10 gigabit Ethernet is becoming more popular in both enterprise and carrier networks, with development starting on 100G Ethernet.

Through Ethernet's history there have also been RF versions of Ethernet, both wireline and wireless. The currently recommended RF wireless networking standards, 802.11 and 802.16, are not Ethernet, in that they do not use the Ethernet link-layer header, and use control and management packet types that don't exist in Ethernet it would not be simply a matter of modulation to transmit Ethernet packets on an 802.11 or 802.16 network, or to transmit 802.11 or 802.16 packets on an Ethernet network.

10BaseT and 100BaseT

We discuss 10BaseT and100BaseT Ethernet together, as they are similar technologies. The most important difference between them is that 10BaseT transmits at

10 Mbps and 100BaseT Ethernet transmits at 100 Mbps. 100BaseT is also commonly called "fast Ethernet" and "100 Mbps Ethernet". 10BaseT and 100BaseT are also very popular Ethernet technologies; in fact, for new installations, 10BaseT and Ethernet are often today the technology of choice. Both 10BaseT and 100BaseT Ethernet use a star topology, as shown in Figure-2.3.

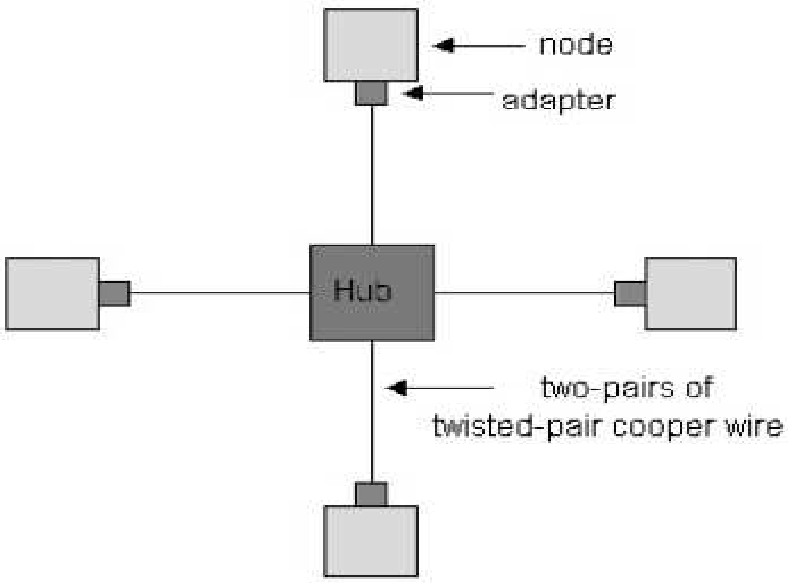


Figure 2.3: Star topology for 10BaseT and 100BaseT

In the star topology there is a central device called a hub (also sometimes called a concentrator.) Each adapter on each node has a direct, point-to-point connection to the hub. This connection consists of two pairs of twisted-pair cooper wire, one for transmitting and the other for receiving. At each end of the connection there is a connector that resembles the RJ-45 connector used for ordinary telephones. The "T" in 10BaseT and 100BaseT stands for "twisted pair". For both 10BaseT and 100BaseT, the maximum length of the connection between an adapter and the hub is 100 meters; the maximum length between any two nodes is 200 meters. As we will discuss in the next section, this maximum distance can be increased by using tiers of hubs, bridges, switches and fiber links. A 10BaseT In essence, a hub is a repeater: when it receives a bit from an adapter, it sends the bit to all the other adapters. In this manner, each adapter can (1) sense the channel to determine if it is idle, and (2) detect a collision while it is transmitting. But hubs are popular because they also provide network management features. For example, if an adapter malfunctions and continually sends Ethernet frames (a so-called "jabbering adapter"), then in a 10Base2 Ethernet will become totally disfunctional; none of the nodes will be able to communicate. But a 10BaseT network will continue to function, because the hub will detect the problem and internally disconnect the malfunctioning adapter. With this feature, the network administrator doesn't have to get out of bed and drive back to work in order to correct the problem for hackers who work late at night. Also, most hubs can gather information and report the information to a host that connects directly to the hub. This monitoring host provides a graphical interface that displays statistics and graphs, such as bandwidth usage, collision rates, average frame sizes, etc. Network administrators can use this information to not only debug and correct problems, but also to plan how the LAN should evolve in the future.

Many Ethernet adapters today are 10/100 Mbps adapters. This means that they can be used for both 10BaseT and 100BaseT Ethernets. 100BaseT, which typically uses category-5 twisted pair (a high quality twisted pair with a lot of twists). Unlike the 10Base2 and 10BaseT, 100BaseT does not use Manchester encoding, but instead a more efficient encoding called 4B5B: every group of five clock periods is used to send 4 bits in order to provide enough transitions to allow clock synchronization.

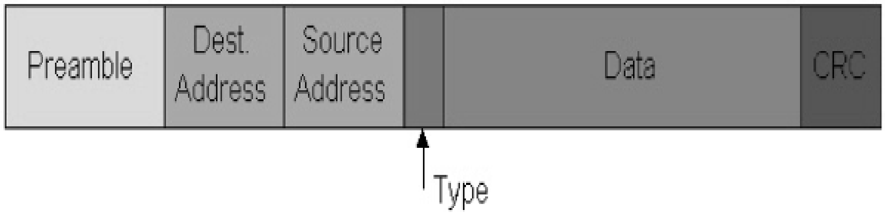


Figure 2.4: Ethernet frame structure

2.3 Token Ring (IEEE 802.5)

The IEEE 802.5 standards specify layers 1 and 2 of a family of *token ring* networks. These networks transmit at 4 Mbps or 6 Mbps. These networks have advantage that, unlike ethernet networks, each node is guaranteed to be allowed to transmit before a specific time. Also, the token ring networks are more efficient than Ethernet networks under high load.

2.3.1 Physical Layer

In a token ring, the nodes are connected into a ring by point-to-point links. (See the Figure-2.5) A network interface has two possible configurations repeater and open. In the repeater configuration, the interface repeats the incoming signal on the outgoing link with a delay of a few bit transmission times. At the same time, the interface copies the signal for the computer. In the open configuration, the interface transmits on the outgoing link and listens on the incoming link. The transmission rate is 4 Mbps or 16 Mbps, as already mentioned. As with Ethernet, signals may be transmitted over a variety of cabling arrangements.

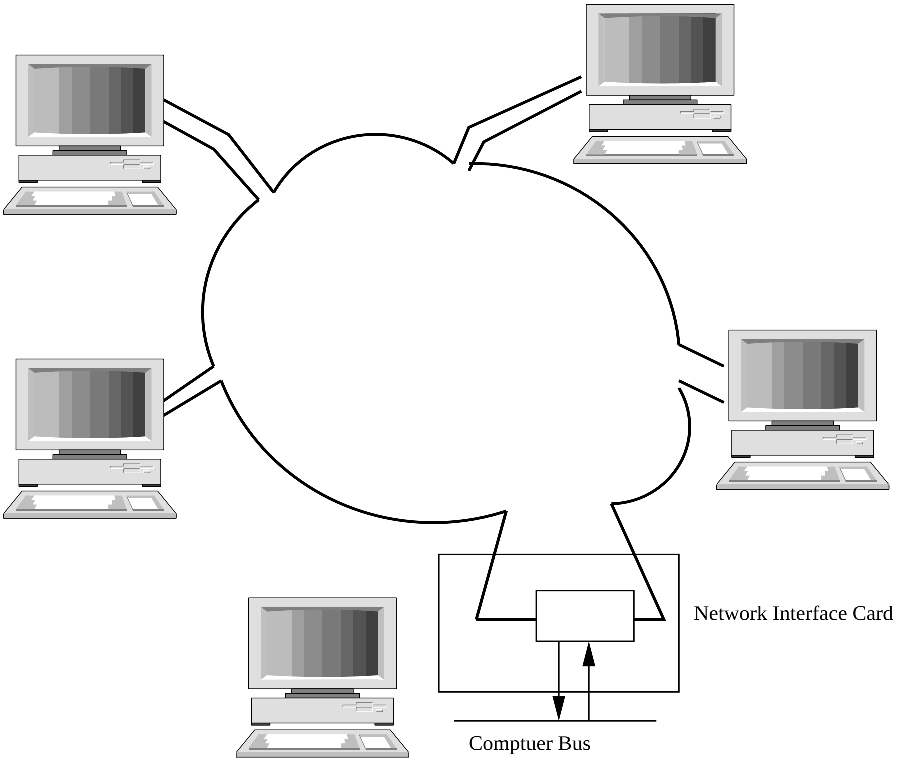


Figure 2.5: Layout of token ring network. The computers are attached by unidirectional point-to-point links around ring

2.3.2 MAC

The frame format is similar to that of the Ethernet packets (Figure-2.4), expect that it used an ending delimiter instead of a length indication. The token is a 3-byte frame that consists of a start frame, an access control, and an ending delimiter, each 1 byte long. The access control field indicates that the 3-byte frame is a token that any station may grab and not a packet.

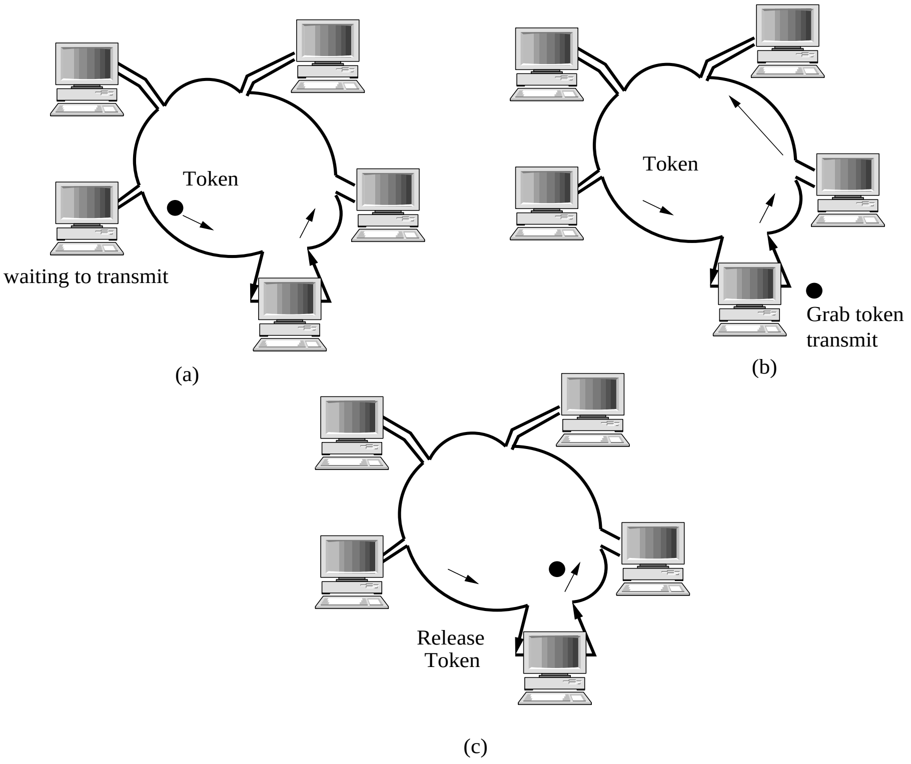


Figure 2.6: Steps in the transmission of a packet when the computers use the release after transmission token-passing protocol

The transmissions proceed in one direction along the ring. Figure-2.6 shows the sequence of events when a node wants to transmit a packet:

- the node waits for the token

- then transmits some of its packets and releases the token

We call this version of the MAC protocol, where a node releases the token right after it finishes transmitting its packets, *release after transmission.* The 16-Mbps token ring networks use this protocol. In a 4-Mbps token ring network, a node that transmits waits until it has completely received uits last packet before releasing the token. We call this version *release after reception.* The standard specifices that a node can hold onto the token and transmit for up to some time, called the *token holding time* (THT), before releasing the token. A typical value of THT is 10 ms.

We use Figure-2.7 to analyze the efficiency of the release after reception protocol. Assume that there are N nodes on a token passing ring. We define T_n to be the time during which node n transmits when it gets the token, before it releases the token. Thus, T_n can range from 0 to THT. We assume that all the nodes want to transmit, so that $T_n > 0$ for $n = 1, \cdots, N$. At time 0, the first node starts transmitting a packet. At time T_1 the first node has transmitted its packets. The last packet has completely returned to the first node $PROP$ seconds later, where $PROP$ is the propagation time of signal around ring. Therefore, at time $T_1 + PROP$ the first node starts transmitting the token, which releases the second node after a propagation time designated by $PROP_{1 \to 2}$. Node 2 then goes through the same sequence of steps node 1 did, and so do the other nodes, one after the another. Eventually, the token comes back to node 1. The efficiency of the token ring is the fraction of time that the nodes transmit packets. The Figure-2.7 shows that the efficiency is approximately equal to $1/(1+a)$ where $a = PROP/E(T_n)$. In this expression, $E(T_n)$ is the average duration of the node transmission. The Figure-2.7 also shows representative values of the efficiency, assuming that the nodes transmit a single packet of the fixed size. As we can see, the efficiency of typical token ring network is more than 90%.

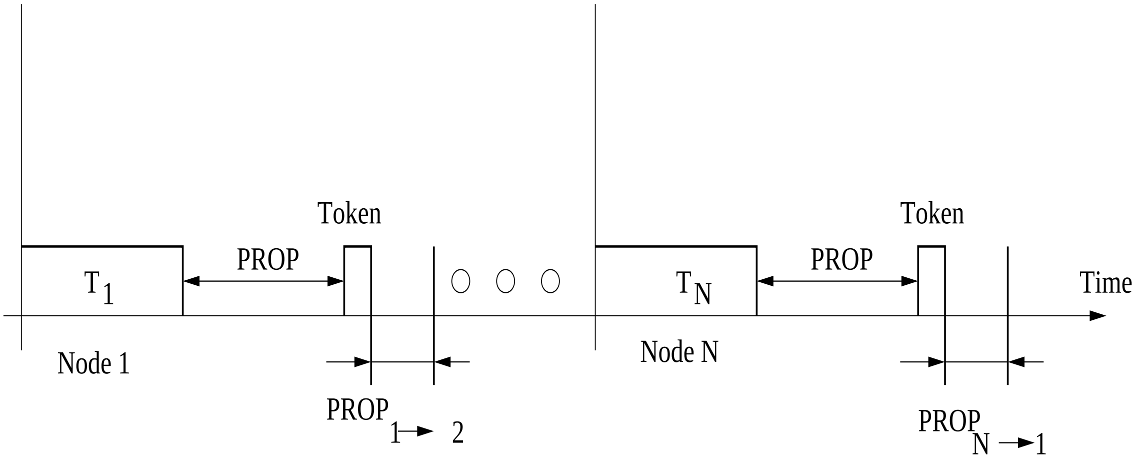


Figure 2.7: Timing diagram for the release after reception token-passing protocol when all the computers have packets to transmit

2.4 Fiber Distributed Data Interface (FDDI)

FThe Fiber Distributed Data Interface (FDDI) specifies a 100-Mbps token-passing, dual-ring LAN using fiber-optic cable. FDDI is frequently used as high-speed backbone technology because of its support for high bandwidth and greater distances than copper. It should be noted that relatively recently, a related copper specification, called Copper Distributed Data Interface (CDDI), has emerged to provide 100-Mbps service over copper. CDDI is the implementation of FDDI protocols over twisted-pair copper wire. This chapter focuses mainly on FDDI specifications and operations, but it also provides a high-level overview of CDDI.

FDDI uses dual-ring architecture with traffic on each ring flowing in opposite directions (called counter-rotating). The dual rings consist of a primary and a secondary ring. During normal operation, the primary ring is used for data transmission, and the secondary ring remains idle. As will be discussed in detail later in this chapter, the primary purpose of the dual rings is to provide superior reliability and robustness. Figure-2.8 shows the counter-rotating primary and secondary FDDI rings.

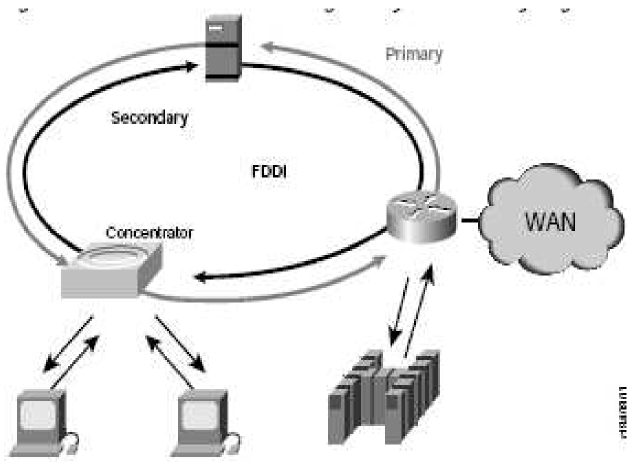


Figure 2.8: FDDI Uses Counter-Rotating Primary and Secondary Rings

2.4.1 FDDI Specifications

FDDI specifies the physical and media-access portions of the OSI reference model. FDDI is not actually a single specification, but it is a collection of four separate specifications, each with a specific function. Combined, these specifications have the capability to provide high-speed connectivity between upper-layer protocols such as TCP/IP and IPX, and media such as fiber-optic cabling. FDDIs four specifications are the Media Access Control (MAC), Physical Layer Protocol (PHY), Physical-Medium Dependent (PMD), and Station Management (SMT) specifications. The MAC specification defines how the medium is accessed, including frame format, token handling, addressing, algorithms for calculating cyclic redundancy check (CRC) value, and error-recovery mechanisms. The PHY specification defines data encoding/decoding procedures, clocking requirements, and framing, among other functions. The PMD specification defines the characteristics of the transmission medium, including fiber-optic links, power levels, bit-error rates, optical components, and connectors. The SMT specification defines FDDI station configuration, ring configuration, and ring control features, including station insertion and removal, initialization, fault isolation and recovery, scheduling, and statistics collection.

FDDI is similar to IEEE 802.3 Ethernet and IEEE 802.5 Token Ring in its relationship with the OSI model. Its primary purpose is to provide connectivity between upper OSI layers of common protocols and the media used to connect network devices. Figure 8-3 illustrates the four FDDI specifications and their relationship to each other and to the IEEE-defined Logical Link Control (LLC) sublayer. The LLC sublayer is a component of Layer 2, the MAC layer, of the OSI reference model.

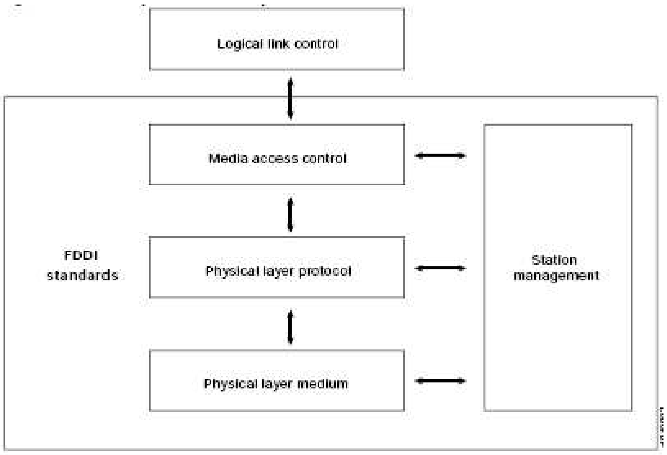


Figure 2.9: FDDI Specifications Map to the OSI Hierarchical Model

2.4.2 FDDI Frame Format

The FDDI frame format is similar to the format of a Token Ring frame. This is one of the areas in which FDDI borrows heavily from earlier LAN technologies, such as Token Ring. FDDI frames can be as large as 4,500 bytes. Figure-2.10 shows the frame format of an FDDI data frame and token.

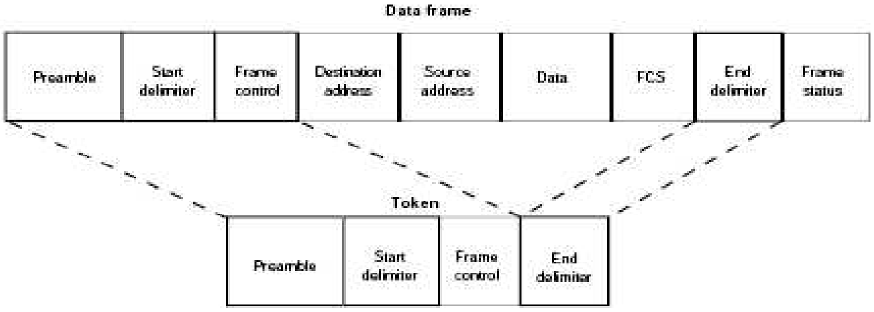


Figure 2.10: The FDDI Frame Is Similar to That of a Token Ring Frame

PreambleGives a unique sequence that prepares each station for an upcoming frame.

- **Start delimiter** -Indicates the beginning of a frame by employing a signaling pattern that differentiates it from the rest of the frame.
- **Frame control** -Indicates the size of the address fields and whether the frame contains asynchronous or synchronous data, among other control information.
- **Destination address** -Contains a unicast (singular), multicast (group), or broadcast (every station) address. As with Ethernet and Token Ring addresses, FDDI destination addresses are 6 bytes long.
- **Source address** -Identifies the single station that sent the frame. As with Ethernet and Token Ring addresses, FDDI source addresses are 6 bytes long.

- **Data** -Contains either information destined for an upper-layer protocol or control information.
- **Frame check sequence (FCS)** -Is filed by the source station with a calculated cyclic redundancy check value dependent on frame contents (as with Token Ring and Ethernet). The destination address recalculates the value to determine whether the frame was damaged in transit. If so, the frame is discarded.
- **End delimiter** -Contains unique symbols; cannot be data symbols that indicate the end of the frame.
- **Frame status** -Allows the source station to determine whether an error occurred; identifies whether the frame was recognized and copied by a receiving station.

2.5 Distributed Queue Dual Bus (DQDB)

The Distributed Queue Dual Bus (DQDB) is a data link layer communication protocol designed for use in metropolitan-area networks (MANs). DQDB specifies a network topology composed of two unidirectional logical buses that interconnect multiple systems. It is defined in the IEEE 802.6 DQDB standard.

An access DQDB describes just the operation of the DQDB protocol (in SMDS, SIP) across a user-network interface (in SMDS, across the SNI). Such operation is distinguished from the operation of a DQDB protocol in any other environment (for example, between carrier equipment within the SMDS PDN).

The access DQDB is composed of the basic SMDS network components:

- **Carrier equipment**- A switch in the SMDS network operates as one station on the bus.
- **Customer Premises Equipment (CPE)**- One or more CPE devices operate as stations on the bus.
- **Subscriber Network Interface (SNI)**- The SNI acts as the interface between the CPE and the carrier equipment.

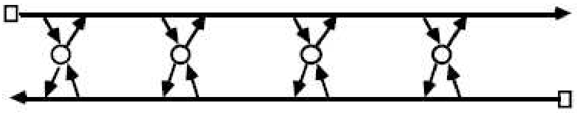


Figure 2.11: Distributed Queue Dual Bus

An SMDS access DQDB typically is arranged in a single-CPE configuration or a multi-CPE configuration. A single-CPE access DQDB configuration consists of one switch in the carrier SMDS network and one CPE station at the subscriber site. Single-CPE DQDB configurations create a two-node DQDB subnetwork. Communication occurs only between the switch and the one CPE device across the SNI. No contention is on the bus because no other CPE devices attempt to access it. A multi-CPE configuration consists of one switch in the carrier SMDS network and a number of interconnected CPE devices at the subscriber site (all belonging to the same subscriber). In multi-CPE configurations, local communication between CPE devices is possible. Some local communication will be visible to the switch serving the SNI, and some will not. Contention for the bus by multiple devices requires the use of the DQDB distributed queuing algorithm, which makes implementing a multi-CPE configuration more complicated than implementing a single-CPE configuration.

The Figure-2.12 shows the topology of DQDB. Each station is attached to two unidirectional buses. The word *bus* is a misnomer, becasue the connections in each direction are implemented by a sequence of point-topoint links instead of abus as in Ethernet or a token bus.

The DQDB MAC protocol is clever way to regulate access to the mdeium as is all stations placed their packets in a single queue that is served on a first-come, first-serve (FCFS) basis. Such a first-come, first-serve protocol would be the fairest possible. However, it cannot be achieved perfectly becasue the queues are distributed in the different stations and no station knows exactly when the other stations got packets to transmit.

A station wanting to transmit to another station situated to its right must use the upper bus, and it must use the lower bus to transmit to stations on its left. THe operations of the two buese are identical; how the nodes transmit on the upper bus is shown in the Figure-2.13.

2.6 Frame Delay

Frame Relay is a high-performance WAN protocol that operates at the physical and data link layers of the OSI reference model. Frame Relay originally was designed for use across Integrated Services Digital Network (ISDN) interfaces. Today, it is used over a variety of other network interfaces as well. This chapter focuses on Frame Relays specifications and applications in the context of WAN services. Frame Relay is an example of a packet-switched technology. Packet-switched networks enable end stations to dynamically share the network medium and the available bandwidth. The following two techniques are used in packet-switching technology:

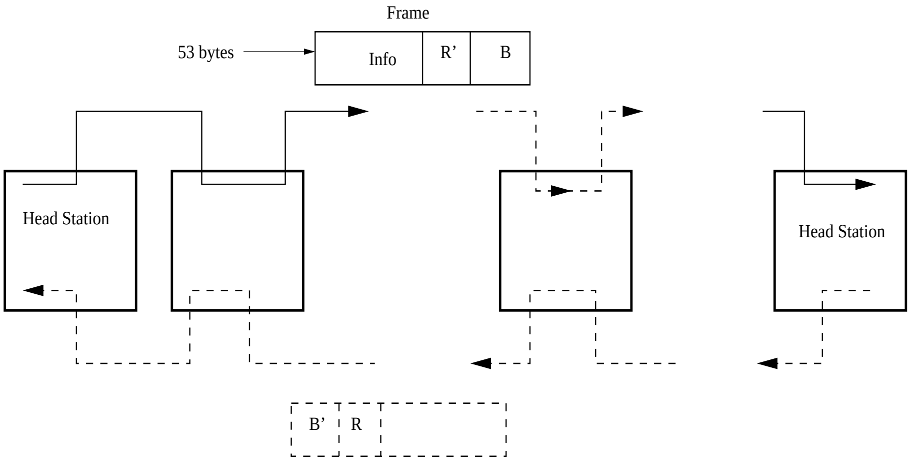
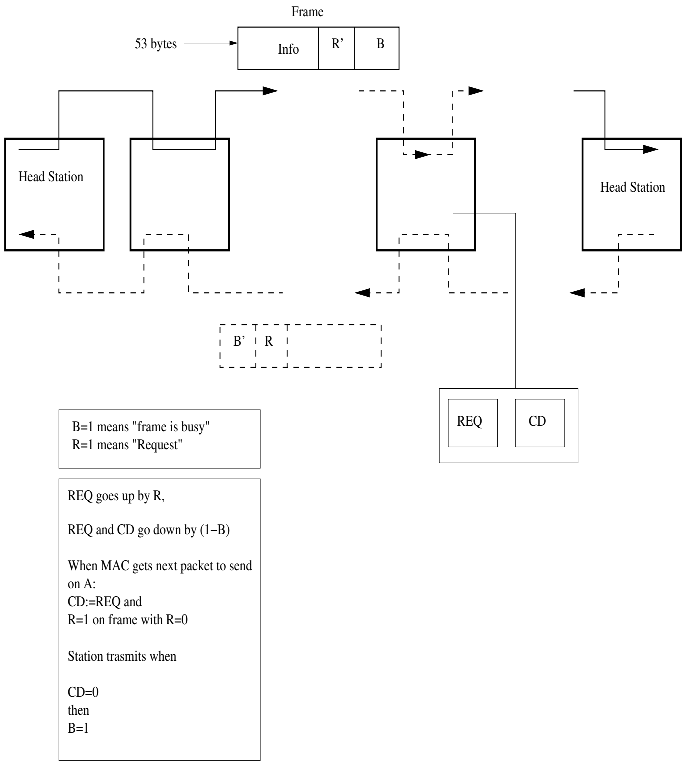


Figure 2.12: In the DQDB network, frames are generated back to back by the left head station with R'=B=0; a station can use a frame if there is no pending request to its right. The operation of the lower bus is similar

- Variable-length packets
- Statistical multiplexing

Variable-length packets are used for more efficient and flexible data transfers. These packets are switched between the various segments in the network until the destination is reached. Statistical multiplexing techniques control network access in a packet-switched network. The advantage of this technique is that it accommodates more flexibility and more efficient use of bandwidth. Most of todays popular LANs, such as Ethernet and Token Ring, are packet-switched networks. Frame Relay often is described as a streamlined version of X.25, offering fewer of the robust capabilities, such as windowing and retransmission of last data that are offered in X.25. This is because Frame Relay typically operates over WAN facilities that offer more reliable connection services and a higher degree of reliability than the facilities available during the late 1970s and early 1980s that served as the common platforms for X.25 WANs. As mentioned earlier, Frame Relay is strictly a Layer 2 protocol suite,



B=1 means "frame is busy"
R=1 means "Request"

REQ goes up by R,

REQ and CD go down by (1−B)

When MAC gets next packet to send
on A:
CD:=REQ and
R=1 on frame with R=0

Station trasmits when

CD=0
then
B=1

Figure 2.13: The figure expalins the operation of the DQDB protocol

whereas X.25 provides services at Layer 3 (the network layer) as well. This enables Frame Relay to offer higher performance and greater transmission efficiency than X.25, and makes Frame Relay suitable for current WAN applications, such as LAN interconnection.

2.6.1 Frame Relay Devices

Devices attached to a Frame Relay WAN fall into the following two general categories:

- Data terminal equipment (DTE)
- Data circuit-terminating equipment (DCE)

DTEs generally are considered to be terminating equipment for a specific network and typically are located on the premises of a customer. In fact, they may be owned by the customer. Examples of DTE devices are terminals, personal computers, routers, and bridges. DCEs are carrier-owned internetworking devices. The purpose of DCE equipment is to provide clocking and switching services in a network, which are the devices that actually transmit data through the WAN. In most cases, these are packet switches. Figure-2.14 shows the relationship between the two categories of devices.

The connection between a DTE device and a DCE device consists of both a physical layer component and a link layer component. The physical component defines the mechanical, electrical, functional, and procedural specifications for the connection between the devices. One of the most commonly used physical layer interface specifications is the recommended standard (RS)-232 specification. The link layer component defines the protocol that establishes the connection between the DTE device, such as a router, and the DCE device, such as a switch.

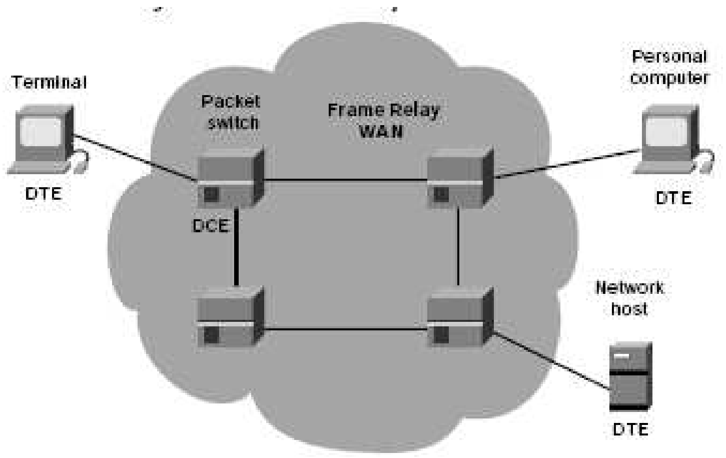


Figure 2.14: DCEs Generally Reside Within Carrier-Operated WANs

2.6.2 Frame Relay Frame Formats

To understand much of the functionality of Frame Relay, it is helpful to understand the structure of the Frame Relay frame. Figure-2.15 depicts the basic format of the Frame Relay frame, and Figure-2.16 illustrates the LMI version of the Frame Relay frame. Flags indicate the beginning and end of the frame. Three primary components make up the Frame Relay frame: the header and address area, the user-data portion, and the frame check sequence (FCS). The address area, which is 2 bytes in length, is comprised of 10 bits representing the actual circuit identifier and 6 bits of fields related to congestion management. This identifier commonly is referred to as the data-link connection identifier (DLCI). Each of these is discussed in the descriptions that follow.

Standard Frame Relay Frame

Standard Frame Relay frames consist of the fields illustrated in Figure-2.15.

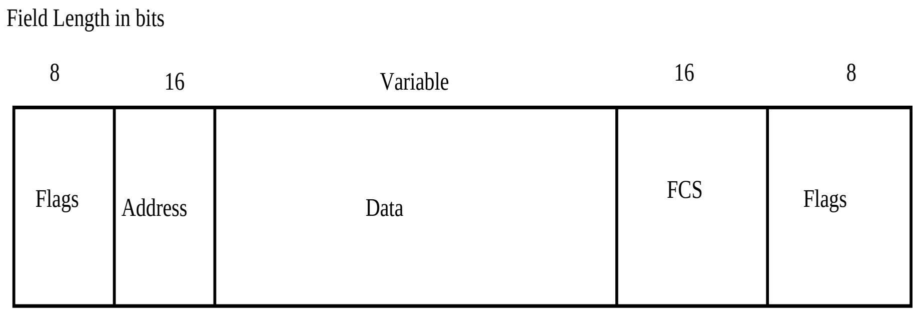


Figure 2.15: Five Fields Comprise the Frame Relay Frame

The following descriptions summarize the basic Frame Relay frame fields illustrated in Figure-2.15.

- Flags
  - Delimits the beginning and end of the frame. The value of this field is always the same and is represented either as the hexadecimal number 7E or as the binary number 01111110.
- Address -Contains the following information:
  - DLCI -The 10-bit DLCI is the essence of the Frame Relay header. This value represents the virtual connection between the DTE device and the switch. Each virtual connection that is multiplexed onto the physical channel will be represented by a unique DLCI. The DLCI values have local significance only, which means that they are unique only to the physical channel on which they reside. Therefore, devices at opposite ends of a connection can use different DLCI values to refer to the same virtual connection.
  - Extended Address (EA) -The EA is used to indicate whether the byte in which the EA value is 1 is the last addressing field. If the value

is 1, then the current byte is determined to be the last DLCI octet. Although current Frame Relay implementations all use a two-octet DLCI, this capability does allow longer DLCIs to be used in the future. The eighth bit of each byte of the Address field is used to indicate the EA.

- Command/Responce (C/R) -The C/R is the bit that follows the most significant DLCI byte in the Address field. The C/R bit is not currently defined.

- Congestion Control-This consists of the 3 bits that control the Frame Relay congestion-notification mechanisms. These are the FECN, BECN, and DE bits, which are the last 3 bits in the Address field.

  Forward-explicit congestion notification (FECN) is a single-bit field that can be set to a value of 1 by a switch to indicate to an end DTE device, such as a router, that congestion was experienced in the direction of the frame transmission from source to destination. The primary benefit of the use of the FECN and BECN fields is the capability of higher-layer protocols to react intelligently to these congestion indicators. Today, DECnet and OSI are the only higher-layer protocols that implement these capabilities.

  Backward-explicit congestion notification (BECN) is a single-bit field that, when set to a value of 1 by a switch, indicates that congestion was experienced in the network in the direction opposite of the frame transmission from source to destination.

  Discard eligibility (DE) is set by the DTE device, such as a router, to indicate that the marked frame is of lesser importance relative to other frames being transmitted. Frames that are marked as "discard eligible" should be discarded before other frames in a congested network. This allows for a basic prioritization mechanism in Frame Relay networks.

- Data -Contains encapsulated upper-layer data. Each frame in this variable-length field includes a user data or payload field that will vary in length up to 16,000 octets. This field serves to transport the higher-layer protocol packet (PDU) through a Frame Relay network.

- Frame Check Sequence -Ensures the integrity of transmitted data. This value is computed by the source device and verified by the receiver to ensure integrity of transmission.

Local Management Interface (LMI) Frame Format

Frame Relay frames that conform to the LMI specifications consist of the fields illustrated in Figure-2.16 The following descriptions summarize the fields illustrated in Figure-2.16.

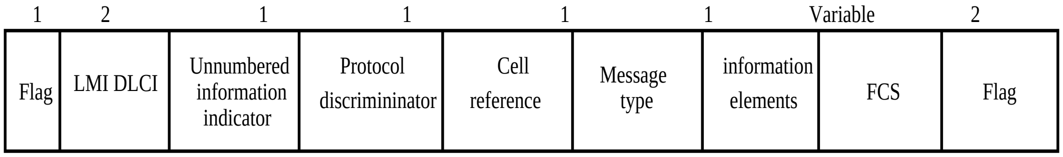


Figure 2.16: Nine Fields Comprise the Frame Relay That Conforms to the LMI Format

- **Flag** -Delimits the beginning and end of the frame.

- **LMI DLCI** -Identifies the frame as an LMI frame instead of a basic Frame Relay frame. The LMI-specific DLCI value defined in the LMI consortium specification is DLCI = 1023.

- **Unnumbered Information Indicator** -Sets the poll/final bit to zero.

- **Protocol Discriminator** -Always contains a value indicating that the frame is an LMI frame.

- **Call Reference** -Always contains zeros. This field currently is not used for any purpose.
- **Message Type** -Labels the frame as one of the following message types:
  - Status-inquiry message-Allows a user device to inquire about the status of the network.
  - Status message-Responds to status-inquiry messages. Status messages include keepalives and Permanent Virtual Circuits (PVC) status messages.
  - Information Elements (IE)-Contains a variable number of individual information elements (IEs). IEs consist of the following fields:
  - IE Identifier-Uniquely identifies the IE.
  - IE Length-Indicates the length of the IE.
  - Data-Consists of 1 or more bytes containing encapsulated upper-layer data.
- Frame Check Sequence (FCS)Ensures the integrity of transmitted data.

2.7 Switched Multimegabit Data Service (SMDS)

Switched Multimegabit Data Service (SMDS) is a high-speed, packet-switched, datagram-based WAN networking technology used for communication over public data networks (PDNs). SMDS can use fiber or copper-based media; it supports speeds of 1.544 Mbps over Digital Signal level 1 (DS-1) transmission facilities, or 44.736 Mbps over Digital Signal level 3 (DS-3) transmission facilities. In addition, SMDS data units are large enough to encapsulate entire IEEE 802.3, IEEE 802.5, and Fiber Distributed Data Interface (FDDI) frames.

2.7.1 SMDS Network Components

SMDS networks consist of several underlying devices to provide high-speed data service. These include customer premises equipment (CPE), carrier equipment, and the subscriber network interface (SNI). CPE is terminal equipment typically owned and maintained by the customer. CPE includes end devices, such as terminals and personal computers, and intermediate nodes, such as routers, modems, and multiplexers. Intermediate nodes, however, sometimes are provided by the SMDS carrier. Carrier equipment generally consists of high-speed WAN switches that must conform to certain network equipment specifications, such as those outlined by Bell Communications Research (Bellcore). These specifications define network operations, the interface between a local carrier network and a long-distance carrier network, and the interface between two switches inside a single carrier network.

The SNI is the interface between CPE and carrier equipment. This interface is the point at which the customer network ends and the carrier network begins. The function of the SNI is to render the technology and operation of the carrier SMDS network transparent to the customer. Figure-2.17 illustrates the relationships among these three components of an SMDS network.

2.7.2 SMDS Interface Protocol

The SMDS Interface Protocol (SIP) is used for communications between CPE and SMDS carrier equipment. SIP provides connectionless service across the subscriber network interface (SNI), allowing the CPE to access the SMDS network. SIP is based on the IEEE 802.6 Distributed Queue Dual Bus (DQDB) standard for cell relay across metropolitan-area networks (MANs). The DQDB was chosen as the basis for SIP because it is an open standard that supports all the SMDS service features. In addition, DQDB was designed for compatibility with current carrier

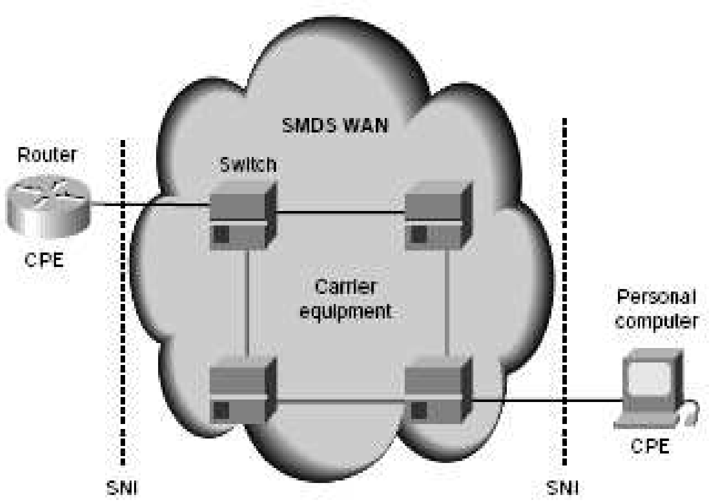


Figure 2.17: The SNI Provides an Interface Between the CPE and the Carrier Equipment in SMDS

transmission standards, and it is aligned with emerging standards for Broadband ISDN (BISDN), which will allow it to interoperate with broadband video and voice services. Figure-2.18 illustrates where SIP is used in an SMDS network.

SIP Levels

SIP consists of three levels. SIP Level 3 operates at the Media Access Control (MAC) sublayer of the data link layer of the OSI reference model. SIP Level 2 operates at the MAC sublayer of the data link layer. SIP Level 1 operates at the physical layer of the OSI reference model. Figure-2.19 illustrates how SIP maps to the OSI reference model, including the IEEE data link sublayers.

SIP Level 3 begins operation when user information is passed to it in the form of SMDS service data units (SDUs). SMDS SDUs then are encapsulated in a SIP

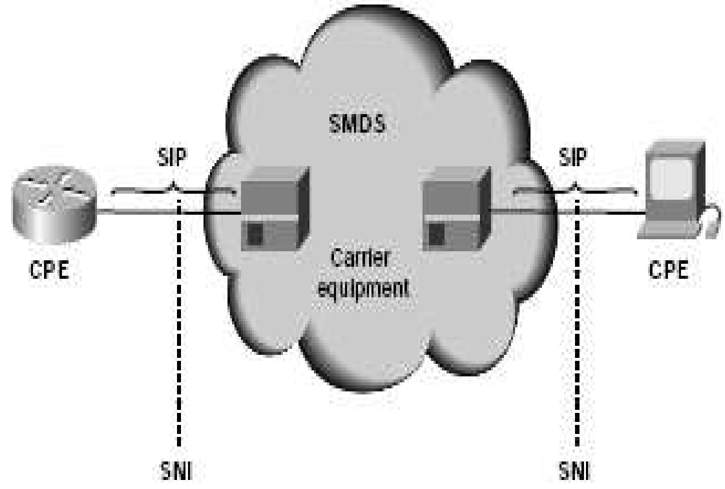


Figure 2.18: SIP Provides Connectionless Service Between the CPE and Carrier Equipment

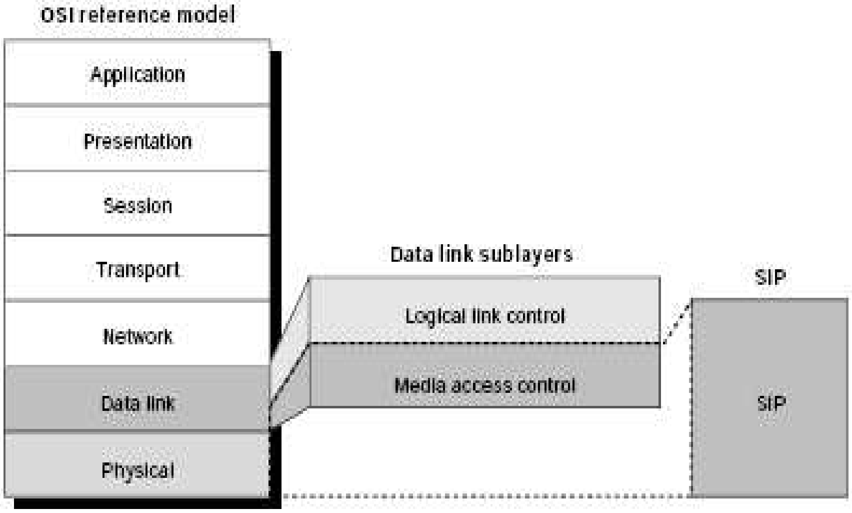


Figure 2.19: SIP Provides Services Associated with the Physical and Data Link Layers of the OSI Model

Level 3 header and trailer. The resulting frame is called a Level 3 protocol data unit (PDU). SIP Level 3 PDUs then are passed to SIP Level 2.

SIP Level 2, which operates at the Media Access Control (MAC) sublayer of the data link layer, begins operating when it receives SIP Level 3 PDUs. The PDUs then are segmented into uniformly sized (53-octet) Level 2 PDUs, called cells. The cells are passed to SIP Level 1 for placement on the physical medium.

SIP Level 1 operates at the physical layer and provides the physical-link protocol that operates at DS-1 or DS-3 rates between CPE devices and the network. SIP Level 1 consists of the transmission system and Physical Layer Convergency Protocol (PLCP) sublayers. The transmission system sublayer defines the characteristics and method of attachment to a DS-1 or DS-3 transmission link. The PLCP specifies how SIP Level 2 cells are to be arranged relative to the DS-1 or DS-3 frame. PLCP also defines other management information.

2.8 Summary

Name	Speed, Connectivity	Delay	Application
Ethernet	10-100-1000 Mbps, local area	Random, increases with load	Transfer of messages between nearby computers
Token Ring	4,16 Mbps, local area	random but bounded	Transfer of messages between nearby computers, some real time traffic
FDDI	100 Mbps, LAN and campus	Random but bounded	LAN interconnections, real-time and CBR applications
DQDB	Unspecified	Random	Unspecified but similar to FDDI
Frame Relay	1.5 Mbps, wide area	Random increases with load	Transfer of messages between distant computers
SMDS	1.5 to more than 45 Mbps	Random, traffic shaping	LAN interconnections, migration to ATM

Table 2.2: Summary of Advances in packet-switched networks