Chapter 3

Computer Communication Networks

Objective of this chapter:

The performance of any distributed system is significantly dependent on its communication network. The performance of the computer network is even more important in high performance distributed systems. If the network is slow or inefficient, the distributed system performance becomes slower than that can be achieved using a single computer. The main objective of this chapter is to briefly review the basic principles of computer network design and then focus on high speed network technologies that will play an important role in the development and deployment of high performance distributed systems.

Key Terms

LAN, MAN, WAN, LPN, Ethernet, CSMA/CD, FDDI, DQDB, ATM, Infiniband, Wireless LAN

3.1 Introduction

Computer networking techniques can be classified based on the transmission medium (fiber, copper, wireless, satellite, etc.), switching technique (packet switching or circuit switching), or distance. The most widely used technique is the one based on distance that classifies computer network into four categories: Local Area Networks (LAN), Metropolitan Area Networks (MAN), Wide Area Networks (WAN), and Local Peripheral Networks (LPN). A LPN covers a relatively short distance (tens or hundreds of meters) and is used mainly to interconnect input/output subsystems to one or more computers. A LAN covers a building or a campus area (few Kilometers) and usually has a simple topology (bus, ring, or star). Most of the current distributed systems are designed using LANs that operate at 10 to 1000 million bits per second (Mbps). A MAN covers a larger distance than a LAN, typically a city or a region (around 100 Kilometers). LANs are usually owned and controlled by one organization, whereas MANs typically use the services of one or more telecommunication providers. A WAN can cover a whole country or one or more continents and it utilizes the network(s) provided by several telecommunications carriers.

In this chapter, we briefly review each computer network type (LAN, MAN, WAN and LPN) and then discuss in detail the high speed network technology associated with each computer network type. The high speed networks to be
discussed in detail include Fiber Distributed Data Interface (FDDI), Distributed Queue Data Buffer (DQDB) and Asynchronous Transfer Mode (ATM) networks. A more detailed description of other types of computer networks can be found in texts that focus mainly on computer networks [Tanenbaum, 1988; Stallings, 1995; Strohl, 1991; Kalmanek, Kanakia and Keshav, 1990].

3.2 LOCAL AREA NETWORKS (LAN)

LANs typically support data transmission rates from 10 Mbps to Giga bit per second (Gbps). Since LANs are designed to span short distances they are able to transmit data at high rates with a very low error rate. The topology of a LAN is usually simple and can be either a bus, ring or a star. The IEEE 802 LAN standards shown in Figure 3.1 define the bottom three layers of the OSI Reference Model. Layer one is the physical layer, layer two is composed of the medium access control (MAC) sub-layer and the logical link control (LLC) sub-layer and layer three is the network layer.

The main IEEE LAN standards include Ethernet, Token Ring and FDDI. Ethernet is until now by far the most widely used LAN technology, whereas FDDI is a mature high speed LAN technology.
3.2.1 ETHERNET

Ethernet (IEEE 802.3) is the name of a popular LAN technology invented at Xerox PARC in the early 1970s. Xerox used an Ethernet-based network to develop a distributed computing system in which users on workstations (clients) communicated with servers of various kinds including file and print servers. The nodes of an Ethernet are connected to the network cable via a transceiver and a tap. The tap and transceiver make the physical and logical connection onto the Ethernet cable. The transceiver contains logic, which controls the transmission and reception of serial data to and from the cable.

Ethernet uses Carrier Sense Multiple Access with Collision (CSMA/CD) protocols to control access to the cable [Bertsekas, 1995; Dalgic, Chien, and Tobagi, 1994].

CSMA is a medium access protocol in which a node listens to the Ethernet medium before transmitting. The probability of the collision is reduced because the transmitting node transmits its message only after it found that the transmission medium is idle. When a node finds out that the medium is inactive, it begins transmitting after waiting a mandatory period to allow the network to settle. However, because of the propagation delay, there is a finite probability that two or more nodes on the Ethernet will simultaneously find the medium in idle state. Consequently, two (or more) transmissions might start at the same time and that result in collisions.

In CSMA/CD, collisions are detected by comparing the transmitted data with the data received from the Ethernet medium to see if the message on the medium matches the message being transmitted. The detection of a collision requires that the collided nodes to retransmit their messages. Retransmission can occur at random or it can follow the exponential back-off rule; an algorithm used for the generation of a random time interval that determines at what time each collided station can transmit its message.

High Speed Ethernet Technologies

Recently, there has been an increased interest in the industry to revive Ethernet by introducing switched fast Ethernet and Gigabit Ethernet.

Fast Ethernet is similar to Ethernet, only ten times faster than Ethernet. Unlike other emerging High Speed network technologies, Ethernet has been installed for over 20 years in business, government, and educational networks. Fast Ethernet uses the same media access protocol (MAC) used in Ethernet (CSMA/CD protocol). This makes the transition from Ethernet to fast Ethernet as well as the inter-networking between Ethernet and fast Ethernet to be a straightforward. Fast Ethernet can work with unshielded twisted-pair cable and thus can be built upon the existing Ethernet wire. That makes fast Ethernet attractive when compared to
other high speed networks such as FDDI and ATM that require fiber optic cables, which will make the upgrade of existing legacy network to such high speed network technologies costly. In designing the fast Ethernet MAC and to make it inter-operate easily with existing Ethernet networks, the duration of each bit transmitted is reduced by a factor of 10. Consequently, this results in increasing the packet speed by 10 times when compared to Ethernet, while packet format and length, error control, and management information remain identical to those of Ethernet. However, the maximum distance between a computer and the fast Ethernet hub/switch depends on the type of cable used and it ranges between 100-400 m.

This High Speed network technology is attractive because it is the same Ethernet technology, but 10 or 100 times faster. Also, fast Ethernet can be deployed as a switching or shared technology as is in Ethernet. However, its scalability and the maximum distance are limiting factors when compared with fiber-based High Speed network technology (e.g., ATM technology).

Gigabit Ethernet (GigE) is a step further from Fast Ethernet. It supports Ethernet speeds of 1Gbps and above and supports Carrier Sense Multiple Access/ Collision Detection (CSMA/CD) as the access method like the Ethernet. It supports both half-duplex and duplex modes of operation. It preserves the existing frame size of 64-1518 bytes specified by IEEE 802.3 for Ethernet. GigE offers speeds compared to ATM at much lesser cost and can support packets belonging to time-sensitive applications in addition to video traffic. The IEEE 802.3z under development will be the standard for GigE. Bit GigE can support a range of 3-5 kms only.

3.2.2 FIBER DISTRIBUTED DATA INTERFACE (FDDI)

FDDI is a high speed LAN proposed in 1982 by the X3T9.5 committee of the American National Standard (ANSI). The X3T9.5 standard for FDDI describes a dual counter-rotating ring LAN that uses a fiber optic medium and a token passing protocol [Tanenbaum, 1988; Ross, 1989; Jain, 1991]. FDDI transmission rate is 100 Mbps. The need for such a High Speed transmission rate has grown from the need for a standard high speed interconnection between computers and their peripherals. FDDI is suitable for front-end networks, typically with an office or a building, which provides interconnection between workstations, file servers, database servers, and low-end computers. Also, the high throughput of FDDI network makes it an ideal network to build a high performance backbone that bridge together several low speed LANs (Ethernet LANs, token rings, token bus).
FDDI uses optical fiber with light emitting diodes (LEDs) transmitting at a nominal wavelength of 1300 nanometers. The total fiber path can be up to 200 kilometers (km) and can connect up to 500 stations separated by a maximum distance of 2 km. An FDDI network consists of stations connected by duplex optical fibers that form dual counter-rotating rings as shown in Figure 3.2. One of the rings is designated as the primary ring and the other one as the secondary. In normal operation data is transmitted on the primary ring. The secondary ring is used as a backup to tolerate a single failure in the cable or in a station. Once a fault is detected, a logical ring is formed using both primary and secondary rings and bypass the faulty segment or station as shown in Figure 3.3.
FDDI Architecture and OSI Model

The FDDI protocol is mainly concerned with only the bottom two layers in the OSI reference model: Physical Layer and Data Link Layer as shown in Figure 3.4. The Physical layer is divided into a Physical (PHY) Protocol sub-layer and the Physical Medium Dependent (PMD) sub-layer. The PMD sub-layer focuses on defining the transmitting and receiving signals, specifying power levels and the types of cables and connectors to be used in FDDI. The physical layer protocol focuses on defining symbols, coding and decoding techniques, clocking requirements, states of the links and data framing formats.

The data Link Layer is subdivided into a Logical Link Control (LLC) sub-layer (the LLC sub-layer is not part of the FDDI protocol specifications) and a Media Access Control (MAC) sub-layer. The MAC sub-layer provides the procedures needed for formatting frames, error checking, token handling and how each station can address and access the network. In addition, a Station Management (SMT) function is included in each layer to provide control and timing for each FDDI station. This includes node configuration, ring initialization, connection and error management. In what follows, we discuss the main functions and algorithms used to implement each sub-layer in the FDDI protocol.

![Figure 3.4 FDDI and the OSI Model](image)

**Physical Medium Dependent (PMD) Sub-Layer**
The Physical Medium Dependent defines the optical hardware interface required to connect to the FDDI rings. This sub-layer deals with the optical characteristics such as the optical transmitters and receivers, the type of connectors to the media, the type of optical fiber cable and an optional bypass optical switch. PMD layer is designed to operate with multimode fiber optics with a wavelength of 1300 nanometer (NM). The distance between two stations is up to 2 kilometers in order to guarantee proper synchronization and allowable data dependent jitter. When a single mode fiber is used, a variation of the PMD sub-layer (SMF-PMD) should be used. The single mode fiber extends the distance between two stations up to 60 kilometers. The PMD document aims to provide link transmission rate with bit error rates (BER) of $2.5 \times 10^{-10}$ at the minimum received power level, and with better than $10^{-12}$ when the power is 2 dB or more above the minimum received power level.

**Physical (PHY) Sub-Layer**

The PHY sub-layer provides the protocols and optical hardware components that support a link from one FDDI station to another. Its main functions are to define the coding, decoding, clocking, and data framing that are required to send and receive data on the fiber medium. The PHY layer supports duplex communications. That is, it provides simultaneous transmission and receiving of data from and to the MAC sub-layer. The PHY sub-layer receives data from MAC sub-layer at the data link layer. It then encodes the data into 4B/5B code format before it is transmitted on the fiber medium as shown in Figure 3.5. Similarly, the receiver receives the encoded data from the medium, determines symbol boundaries based on the recognition of a Start Delimiter, and then forwards decoded symbols to MAC sub-layer.

Data transmitted on the fiber is encoded in a 4 of 5 group code (4B/5B scheme) with each group is referred to as a symbol as shown in Figure 3.5. With 5-bit symbols, there are 32 possible symbols: 16 data symbols, each representing 4 bits of ordered binary data; 3 for starting and ending delimiters, 2 are used as control indicators, and 3 are used for line-state signaling which is recognized by the physical layer hardware. The remaining 8 symbols are not used since they violate code run length and DC balance requirements [Stallings, 1983]. The 4B/5B scheme is relatively efficient on bandwidth since 100 Mbps data rate is transmitted on the fiber optic at 125 Mbps rate. This is better than the Manchester encoding scheme used in Ethernet; a 10 Mbps data transmission rate is transmitted on the medium at 20 Mbps. After the data is encoded into 4B/5B symbols, it is also translated using a Non-Return to Zero Inverted (NRZI) code before it is sent to the optical fiber. The NRZI coding scheme reduces the number of transitions of the transmitted data streams and thus reduces the complexity of FDDI hardware components [Black-Emerging]. In NRZI coding scheme, instead of sending a zero bit as a logic level low (absence of light for optical medium), a zero is transmitted as the absence of a transition from low to high or from high to low. A one is transmitted as a transition from low to high or high to low. The advantage of this technique is that it eliminates the need for defining a threshold level. A pre-defined threshold is susceptible to a drift in the average bias
of the signal. The disadvantage of the NRZI encoding is the loss of the self-clocking property as in Manchester encoding. To compensate for this loss, a long preamble is used to synchronize the receiver to the sender's clock.

<table>
<thead>
<tr>
<th>Code Group</th>
<th>Symbol Assignment</th>
<th>Code Group</th>
<th>Symbol Assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>11110</td>
<td>0000 0</td>
<td>00000</td>
<td>Q</td>
</tr>
<tr>
<td>01000</td>
<td>0001 1</td>
<td>11111</td>
<td>I</td>
</tr>
<tr>
<td>01010</td>
<td>0010 2</td>
<td>00100</td>
<td>H</td>
</tr>
<tr>
<td>10101</td>
<td>0011 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>01010</td>
<td>0100 4</td>
<td>11000</td>
<td>J</td>
</tr>
<tr>
<td>01011</td>
<td>0101 5</td>
<td>10001</td>
<td>K</td>
</tr>
<tr>
<td>01110</td>
<td>0110 6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>01111</td>
<td>0111 7</td>
<td>01101</td>
<td>T</td>
</tr>
<tr>
<td>10010</td>
<td>1000 8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10011</td>
<td>1001 9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10110</td>
<td>1010 A</td>
<td>00111</td>
<td>R</td>
</tr>
<tr>
<td>10111</td>
<td>1011 B</td>
<td>11001</td>
<td>S</td>
</tr>
<tr>
<td>11010</td>
<td>1100 C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11011</td>
<td>1101 D</td>
<td>00001</td>
<td>V or H</td>
</tr>
<tr>
<td>11100</td>
<td>1110 E</td>
<td>00010</td>
<td>V or H</td>
</tr>
<tr>
<td>11101</td>
<td>1111 F</td>
<td>00011</td>
<td>V</td>
</tr>
<tr>
<td></td>
<td></td>
<td>00101</td>
<td>V</td>
</tr>
<tr>
<td></td>
<td></td>
<td>00110</td>
<td>V</td>
</tr>
<tr>
<td></td>
<td></td>
<td>01000</td>
<td>V or H</td>
</tr>
<tr>
<td></td>
<td></td>
<td>01100</td>
<td>V</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10000</td>
<td>V or H</td>
</tr>
</tbody>
</table>

Figure 3.5 FDDI 4B/5B Code Scheme

The clocking method used in FDDI is point-to-point; all stations transmit using their local clocks. The receiving stations decodes the received data by recognizing that a bit 1 will be received when the current bit is the complement of the previous bit and a bit 0 when the current bit is the same as the previous bit. By detecting the transitions in the received data, the receiver station can synchronize its local clock with the transmitter clock. An elasticity buffer (EB) function is used to adjust the slight frequency difference between the recovered clock and the local station clock. The elasticity buffer is inserted between the receiver, which supports a variable frequency clock to track the clock of the previous transmitting station, and the transmitter at the receiver side, which runs on a fixed frequency clock. The elasticity buffer in each station is reinitialized during the preamble (PA), which precedes each frame or token. The transmitter clock has been chosen with 0.005% stability. With an elasticity buffer of 10 bits, frames of up to 4500 bytes in length can be supported without exceeding the limit of the elasticity buffer.

Media Access Control (MAC) sub-layer

The MAC protocol is a timed token ring protocol similar to the IEEE standard 802.5. The MAC sub-layer controls the transmission of data frames on the ring. The formats of the data and token frames are shown in Figure 3.6. The preamble
field is a string of 16 or more non-data symbols that are used to re-synchronize the receiver's clock to the received frame. The frame control field contains information such as to whether the frame is synchronous/asynchronous and whether 16 or 48 bit addresses are used. The ring network must support 16 and 48 bit addresses as well as a global broadcast feature to all stations. The frame check field is a 32-bit cyclic redundancy check (CRC) for the fields. The frame status indicates whether the frame was copied successfully, an error was detected and/or address was recognized. It is used by the source station to determine successful completion of the transmission.

The basic concept of a ring is that each station repeats the frame it receives to its next station [Ross, 1986; Stalling, 1983]. If the destination station address (DA) of the frame matches the MAC’s address, then the frame is copied into a local buffer and the LLC is notified of the frame’s arrival. MAC marks the Frame Status (FS) field to indicate three possible outcomes: 1) Successful recognition of the frame address, 2) the copying of the frame into a local buffer, or 3) the deletion of an erroneous frame. The frame propagates around the ring until it reaches the station that originally placed it on the ring. The transmitting station examines the FS field to determine the success of the transmission. The transmitting station is responsible for removing from the ring all its transmitted frames; this process is referred to as frame stripping. During the stripping phase, the transmitting station inserts IDLE symbols on the ring.

If a station has a frame to transmit then it can do so only after the token has been captured. A token is a special frame, which indicates that the medium is available for use as shown in Figure 3.3. FDDI protocol supports multiple priority levels to assume the proper handling of frames. If the priority of a station does not allow it to capture a token (its priority is less than the priority of the token), it must repeat it to the next station. When a station captures the token, it removes it from the ring, transmits one or more frames depending on the Token Rotation Time (TRT) and Target Token Rotation Time (TTRT) as will be discussed later, and when it is completed, it issues a new token. The new token indicates the availability of the medium for transmission by another station.
Timed Token Protocol

FDDI protocol is a timed token protocol that allows a station to have a longer period for transmission when previous stations do not hold the token too long; they do not have any data frame to send and thus they relinquish the token immediately. During the initialization, a target token rotation time (TTRT) is negotiated, and the agreed value is stored in each station. The actual token rotation time (TRT) is stored in each station and resets once the token arrives. The amount of traffic, both synchronous and asynchronous, that FDDI allows on the network is related to the following equation:

\[ TTRT \geq TRT + THT \]

where \( TRT \) denotes the token rotation time, that is the time since the token was last received; \( THT \) denotes the token holding time, that is the time that the station has held onto the token; and \( TTRT \) denotes the target token rotation time, that is the desired average for the token rotation time.

Essentially, this equation states that on average, the token must circulate around the ring within a pre-determined amount of time. This property explains why FDDI protocol is known as "timed token protocol". The TTRT is negotiated and agreed upon by all the stations at initialization of the network. The determination of TTRT to obtain the best performance has been the subject of many papers and is mainly determined by the desired efficiency of the network, the desired latency in accessing the network, and the expected load on the network [Agrawal, Chen, and Zhao, 1993]. TRT is constantly re-calculated by each station and is equal to the amount of time since the token was last received. THT is the amount of time that a station has held onto the token. A station that has the token and wants to transmit a message must follow the following two rules:

1) Transmit any synchronous frames that are required to be transmitted. Asynchronous frame may be transmitted if \( TTRT \geq TRT + THT \), before the token is released and put back on the ring.

Synchronous traffic has priority over asynchronous traffic because of the deadlines that need to be met. In order to reserve bandwidth for asynchronous traffic, the amount of synchronous traffic allocated to each station is negotiated and agreed upon at network initialization. In addition, FDDI has an asynchronous priority scheme with up to 8 levels based upon the following inequality:

\[ Ti \geq TRT + THT \]

where \( Ti \) denotes the time allocated to transmit asynchronous traffic of priority \( i \) (\( i \) priority can range from 1 to 8). FDDI also contains a multi-frame asynchronous mode, which supports a continuous dialogue between stations. Two
Logical Link Control (LLC) Layer

The Logical Link Control Layer is the means by which FDDI communicates with higher level protocols. FDDI does not define an LLC sub-layer but has been designed to be compatible with the standard IEEE 802.2 LLC format.

Station Management (SMT) Functions

The Station Management Function monitors all the activities on the FDDI ring and provide control over all stations functions. The main functions of the SMT include [Ross, 1986]:

Fault Detection/Recovery

The FDDI protocol contains several techniques to detect, isolate and recover from network failures. The recovery mechanisms can be grouped into two categories: protocol related failures and physical failures.

Connection Management

This involves controlling the bypass switch in each FDDI node, initializing valid PHY links, and positioning MACs on the appropriate FDDI ring.

Frame Handling

This function assists in network configuration. SMT uses a special frame, Next Station Address (NSA) frame, to configure the nodes on the FDDI rings.

Synchronous Bandwidth Management

The highest priority in FDDI is given to synchronous traffic where fixed units of data are to be delivered at regular time intervals. Delivery is guaranteed with a delay not exceeding twice TTRT. The bandwidth required for synchronous traffic is assigned first and the remaining bandwidth is allocated for the asynchronous traffic. In what follows, we discuss the protocol used to initialize the TTRT interval.
For proper operation of FDDI's timed token protocol, every station must agree upon the value of the targeted token rotation time. This initialization of the network is accomplished through the use of claim frames. If a station wants to change the value of the TTRT, it begins to transmit claim frames with a new value of TTRT. Each station that receives a claim frame must do one of two things:

1) If the value of TTRT in the claim frame is smaller than its current value, then use the new TTRT and relay the claim frame.  
2) If the value of TTRT in the claim frame is greater than its current value, then transmit a new claim frame with the smaller TTRT.

Initialization is complete when a station has received its own claim frame. This means that all stations now have the same value of TTRT. The station that received its own claim frame is now responsible for initializing the network to an operational state. FDDI protocol guarantees that the maximum delay that can be incurred in transmitting synchronous traffic is double the value of TTRT. Consequently, if a station needs the delay to be less than an upper bound (DELAYMAX), it attempts to set the TTRT to be equal to half of this upper bound, i.e., TTRT = DELAYMAX /2.

A station can be connected to one or both rings and that connectivity determines the type of protocol functions to be supported by each station. There are three basic types of stations: Dual Attachment Station (DAS), Concentrator, and Single Attachment Station (SAS). The DAS type requires two duplex cables, one to each of the adjacent stations. The concentrator is a special DAS that provides connection to the ring for several low-end Single Attachment Stations. In this case, an SAS node is connected only to one ring and as a result, fault tolerance can be supported with SAS nodes.

**FDDI-II Architecture and OSI Model**

One main limitation of the FDDI synchronous protocol is that although on average frames will reach their destination at a periodic rate defined by TTRT, there is a possibility that a frame may reach its destination with an elapsed time greater than TTRT. This will occur under heavy network loading. For example, assume that one station is required to send some synchronous traffic and when it receives the token, the TRT is equal to TTRT. In this case, no asynchronous frames can be sent but it is still required to transmit its synchronous frame. As a result, the token's TRT will be greater than TTRT and this condition may cause a glitch in an audio or video signal, which must be transmitted at a periodic rate. This limitation of the FDDI's synchronous protocol has led to the development of FDDI-II.

FDDI-II adds to FDDI circuit switching capability so that it can handle the integration of voice, video and data over an FDDI network. In FDDI-II, the
bandwidth is allocated to circuit-switched data in multiples of 6.144 Mbps isochronous channels. The term isochronous refers to the essential characteristics of a time scale or a signal such that the time intervals between consecutive significant instants either have the same duration or duration's that are multiples of the shortest duration [Teener, 1989]. The number of isochronous channels can be up to 16 channels using a maximum of 98.304 Mbps, where each channel can be flexibly allocated to form a variety of highways whose bandwidths are multiple of 8 Kbps (e.g., 8 Kbps, 64 Kbps, 1.53 Mbps or 2.04 Mbps). Consequently, the synchronous and asynchronous traffic may have only 1.024 Mbps bandwidth when the all the 16 isochronous channels are allocated. Isochronous channels may be dynamically assigned and de-assigned on a real-time basis with any of the unassigned bandwidth is allocated to the normal FDDI traffic (synchronous and asynchronous).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>FDDI ANSI X3 t 9.5</th>
<th>Token Ring IEEE 802. 5</th>
<th>Ethernet IEEE 802. 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Rate</td>
<td>100 Mbps</td>
<td>4 or 16 Mbps</td>
<td>10 Mbps</td>
</tr>
<tr>
<td>Overall Length</td>
<td>100 Km</td>
<td>1.2 Km</td>
<td>2.5 Km</td>
</tr>
<tr>
<td>Nodes</td>
<td>500</td>
<td>96</td>
<td>1024</td>
</tr>
<tr>
<td>Distance between Nodes</td>
<td>2 Km</td>
<td>0.46 Km</td>
<td>0.5 Km</td>
</tr>
<tr>
<td>Packet Size (max)</td>
<td>4500 Octets</td>
<td>8191 Octets</td>
<td>1514 Octets</td>
</tr>
<tr>
<td>Medium</td>
<td>Fiber</td>
<td>Twisted Pair / Fiber</td>
<td>Coaxial Cable</td>
</tr>
<tr>
<td>Medium Access</td>
<td>Dual-ring token passing</td>
<td>Single-ring Token passing</td>
<td>CSMA / CD</td>
</tr>
</tbody>
</table>

Table 3.1: Comparison of Ethernet, FDDI and Token Ring

FDDI-II represents a modification to the original FDDI specification such that an additional sub-layer (Hybrid Ring Controller - HRC) has been added to the Data Link Layer. The HRC allows FDDI-II to operate in an upwardly compatible hybrid mode that not only provides the standard FDDI packet transmission capability but it also provides an isochronous transport mode. The function of the HRC is to multiplex the data packets. This divides the FDDI-II data stream into multiple data streams, one for each of the wide band channels that has been allocated. More detailed information about FDDI-II circuit switched data format and how bandwidth is allocated dynamically to isochronous, synchronous and asynchronous traffic can be found in [Ross, 1986].

Copper FDDI (CDDI)
Another cost-effective alternative to fiber optic FDDI is another standard that replaces the fiber optic cables by copper. The 100 Mbps copper FDDI (CDDI) standard would use the same protocol as FDDI except that its transmission medium would use the commonplace unshielded twisted-pair or shielded twisted-pair copper wiring. The main advantage of using copper is that copper wiring, connectors, and transceivers are much cheaper. The main tradeoff in using copper wiring is that the maximum distance that could be traversed between nodes would be limited to possibly 50 or 100 meters before electromagnetic interference becomes a problem. This maximum distance is not a severe limiting factor since the CDDI network would be used mainly for communication within a small LAN that is physically located in one room or laboratory. The CDDI network could then interface to the larger FDDI network through a concentrator station. In this case, the FDDI network acts as a backbone network spanning large distances interconnecting smaller CDDI LANs with a great savings in cost.

3.3 Metropolitan Area Networks

The DQDB is emerging as one of the leading technologies for high-speed metropolitan area networks. DQDB is a media access control (MAC) protocol, which is being standardized as the IEEE 802.6 standard for MANs [Stallings, 1995]. DQDB consists of two 150 Mbps contra-directional buses with two head nodes, one on each bus, that continuously send fixed-length time slots (53 octets) down the buses. The transmission on the two buses is independent and hence the aggregate bandwidth of the DQDB network is twice the data rate of the bus. The clock period of DQDB network is equal to 125 microseconds that has been chosen to support isochronous services, that is voice services that require 8Khz frequency.

The DQDB protocol is divided into three layers: the first layer from the bottom corresponds to the physical layer of the OSI reference mode, the second layer corresponds to the medium access sublayer, and the third layer corresponds to the data-link layer as shown in Figure 3.7. DQDB protocols support three types of services: connection-less, connection-oriented and isochronous services. The main task of the convergence sublayer within a DQDB network is to map user services into the underlying medium-access service. The connection-less service transmits frames of length up to 9188 octets. Using fixed length slots of 52 octets, DQDB provides the capability to perform frame segmentation and re-assembly. The connection-oriented service supports the transmission of 52-octet segments between nodes interconnected by a virtual channel connection. The isochronous service provides similar service to the connection-oriented service, but for users that require a constant inter-arrival time.

DQDB MAC Protocol

DQDB standard specifies the Medium Access Control and the physical layers. Each bus independently transfers MAC cycle frames of duration 125
microseconds, each frame contains a number of short and fixed slots and frame header. The frames are generated by the head node at each bus and flow downstream passing each node before being discarded at the end of the bus. There are two types of slots: Queued Arbitrated Slots (QA) and Pre-Arbitrated Slots (PA). QA slots are used to transfer asynchronous segments and PA slots are used to transfer isochronous segments. In what follows, we focus on how the distributed queue algorithm controls the access to the QA slots.

![Figure 3.7: Functional Block Diagram of a DQDB Node](image)

The DQDB MAC protocol acts like a single first-in-first-out (FIFO) queue. At any given time, the node associated with the request at the top of the queue is allowed to transmit in the first idle slot on the required bus. However, this single queue does not physically exist, but instead it is implemented in a distributed manner using the queues available in each node. This is can be explained as follows.

Each head of a bus continuously generates slots that contain in its header a *BUSY* (BSY) bit and a *REQUEST* (REQ) bit. The busy bit indicates whether or not a segment occupies the slot, while the REQ bit is used for sending requests for future segment transmission. The nodes on each bus counts the slots that have the request bit set and the idle slots that pass by, so that they can determine their position in the global distributed queue and consequently determine when they can start transmitting their data.
Several studies [Stallings, 1987] have shown that the DQDB MAC access protocol is not fair because the node waiting time depends on its position with respect to the slot generators. As a result, several changes have been proposed to make DQDB protocol more fair [22]. Later, we discuss one approach, Bandwidth Balancing Mechanism (BWB) to address the unfairness issue in DQDB.

The DQDB access mechanism associated with one bus can be implemented using two queue buffers and two counters. Without loss of generality, we name the bus A the *forward* bus and bus B the *reverse* bus. We will focus our description on segment transmission in the forward bus, the procedure for transmission in the reverse bus being the same. To implement the DQDB access mechanism on the forward bus, each node contains two counters - Request counter (RC) and Down counter (DC) and two queues, one for each bus. Each node can be in one of two states: idle when there is no segment to transmit, or count down.

**Idle State:** When a node is in the idle state, the node keeps count of the outstanding requests from its downstream nodes using the RC counter. The RC counter increases by one for each request received in the reverse bus and decreased by one for each empty slot in the forward bus; each empty slot on the forward bus will be used to transmit one segment by downstream nodes. Hence, the value of the Request counter (RC) reflects the number of outstanding requests that have been reserved by the downstream nodes.

![Figure 3.8 DQDB MAC Protocol Implementation](image)

**Count Down State:** When the node becomes active and has a segment to transmit, the node transfers the RC counter to the CD counter and resets the RC counter to zero. The node then sends a request in the reverse bus by setting REQ to 1 in the first slot with REQ bit equals to zero. The CD counter is decreased by one for every empty slot on the forward bus until it reaches zero. Immediately after this event, the node transmits into the first empty slot in the forward bus.

**Priority Levels**
DQDB supports three levels of priorities that can be implemented by using separate distributed queues, and two counters for each priority level. This means that each node will have six Request Counters and Down Counters, two for each priority. Furthermore, the segment format will have three request bits, one for each priority level. In this case, a node that wants to transmit on the Bus A with a priority level, say, it will set the request bit corresponding to this priority level in the first slot on Bus B that has not set the bit corresponding to priority. In this case, the Down Counter (DC) is decremented with every free slot passing on Bus A, but is incremented for every request on Bus B with a higher priority than the counter priority. The Request Counter (RC) is incremented only when a passing request slot has the same priority level; the higher priority requests have already been accounted for in the Down Counter [Stallings, 1987].

**DQDB Fairness**

Several research results have shown that DQDB is unfair and the DQDB unfairness depends on the medium capacity and the bus length [Conti, 1991]. The unfairness in DQDB can result in unpredictable behavior at heavy load. One approach to improve the fairness of DQDB is to use the bandwidth balancing mechanism (BWB). In this mechanism, whenever the CD counter reaches a zero and the station transmits in the next empty slot, it sends a signal to the bandwidth balancing machine (BWB). The BWB machine uses a counter to count the number of segments transmitted by its station. Once this counter reaches a given threshold, referred to as BWB-MOD, the counter is cleared and the RQ-CTR is incremented by one. That means, that this station will skip one empty slot to be used by other downstream stations, which are further away from the slot generator on the forward bus and thus improving DQDB fairness. The value of BWB-MOD can vary from 0-16 where the 0 value means the bandwidth balancing mechanism is disabled [conti, 1991].

**Discussion**

A MAN is optimized for a larger geographical area than a LAN, ranging from several blocks of buildings to entire cities. As with local area networks, MANs can also depend on communication channels of moderate-to-high data rates. IEEE 802.6 is an important standard to cover this type of networks as well as LANs. It offers several transmission rates that can initially start at 44.7 Mbps and later expand it to speeds ranging from 1.544 Mbps to 155 Mbps. DQDB is different from FDDI and token ring networks because it uses a high speed shared medium that supports three types of traffic: bursty, asynchronous and synchronous. Furthermore, the use of fixed-length packets, that are compatible with ATM, provides an efficient and effective support for small and large packets and for isochronous data.

3.4 Wide Area Networks (WANs)
The trend for transmission of information generated from facsimile, video, electronic mail, data, and images has speeded up the conversion from analog-based systems to high-speed digital networks. The Integrated Services Digital Network (ISDN) has been recommended as a wide area network standard by CCITT that is expected to handle a wide range of services that cover future applications of high-speed networks. There are two types of ISDN: Narrowband ISDN (N-ISDN) and Broadband ISDN (B-ISDN). The main goal of N-ISDN is to integrate the various services that include voice, video and data. The B-ISDN supports high data rates (hundreds of Mbps). In this section, we discuss the architecture and the services offered by these two types of networks.

3.4.1 Narrowband ISDN (N-ISDN)

The CCITT standard defines an ISDN network as a network that provides end-to-end digital connectivity to support voice and non-voice services (data, images, facsimile, etc.). The network architecture recommendations for ISDN should support several types of networks: packet switching, circuit switching, non-switched, and common-channel signaling. ISDN can be viewed as a digital bit pipe in which multiple sources are multiplexed into this digital pipe. There are several communication channels that can be multiplexed over this pipe and are as follows:

- **B channel**: It operates at 64 Kbps rate and it is used to provide circuit switched, packet switched and semi-permanent circuit interconnections. It is used to carry digital data, digitized voice and mixtures of lower-rate digital traffic.

- **D channel**: It operates at 16 Kbps and it is used for two purposes: for signaling purposes in conjunction with circuit-switched calls on associated B channels, and as a pipe to carry packet-switched or slow-speed telemetry information. For the H channels, three hybrid channel speeds are identified: H0 channel that operates at 384 Kbps, H11 channel that operates at 1.536 Kbps, and H12 channel that operates at 1.92 Kbps. These channels are used for providing higher bit rates for applications such as fast facsimile, high-speed data, high-quality audio and video.

Two combinations of these channels have been standardized: basic access rate and primary access rate. The basic access consists of 2B+D channels, providing 192Kbps (including 48Kbps overhead). Typical applications which use this access mode are those addressing most of the individual users including homes and small offices, like simultaneous use of voice and data applications, teletext, facsimile etc. These services could either use a one multifunctional terminal or several terminals. Usually a single physical link is used for this access mode. The customer can use all or parts of the two B channels and the D channel. Most present day twisted pair loops will support this mode. The primary access mode is
intended for higher data rate communication requirements, which typically fall under the category of nB+D channels. In this mode, the user can use all or part of the B channels and the D channel. This primary access rate service is provided using time division multiplexed signals over four-wire copper circuits or other media. Each B channel can be switched independently; some B channels may be permanently connected depending on the service application. The H channels can also be considered to fall into this category.

Network Architecture and Channels

ISDN Reference Model

ISDN provides users with full network support by adopting the seven layers of the OSI reference model. However, ISDN services are confined to the bottom three layers (physical, data and network layers) of the OSI reference model. Consequently, ISDN offers three main services [stallings, 1993]: Bearer Services, Teleservices, and Supplementary Services. Bearer Services offer information transfer without alteration in real time. This service corresponds to the OSI's network service layer. There are various types of bearer services depending on the type of application sought. Typical applications include speech and audio information transfer.

Teleservices combine the data transportation (using bearer services) and information processing. These services can be considered to be more user friendly services and use terminal equipment. Typical applications are telephony, telefaxes and other computer to computer applications. These correspond to all the services offered by the several layers of the OSI reference model. Supplementary Services are a mixture of one or more bearer or teleservices for providing enhanced services which include Direct-Dial-in, Conference Calling and Credit-card Calling. A detailed description of these services can be found in [stallings, 1987].

- **Physical layer:** This layer defines two types of interfaces depending on the type of access namely Basic interface (basic access) and Primary interface (primary access).
- **Data link layer:** This layer has different link-access protocols (LAP) depending on the channel used for the link, namely LAP-B (balanced for B channel) and LAP-D (for D channel). Apart from these link access techniques, frame-relay access is also a part of the protocol definition.
- **Network layer:** This layer includes separate protocols for packet switching (X.25), circuit switching, semi-permanent connection and channel signaling.

ISDN User-Network Interfaces
A key aspect of ISDN is that a small set of compatible user-network interfaces can support a wide range of user applications, equipment and configurations. The number of user-network interfaces are kept small to maximize user flexibility and to reduce cost. To achieve this goal ISDN standards define a reference model showing the functional groups and reference points between the groups.

Functional groups are sets of functions needed in ISDN user access arrangements. Specific functions in the functional groups may be done in one or multiple pieces of actual equipment. Reference points are conceptual points for dividing the functional groups. In specific implementations, reference points may in fact represent a physical interface between two functional groups. The functional groups can be classified into two types of devices: Network Termination (NT1, NT2 and NT12) and Terminal Equipment (TE1 and TE2). Network Termination 1 (NT1) provides functions similar to those offered by the physical layer of the OSI Reference Model. Network Termination 2 (NT2) provides functions equivalent to those offered by layers 1 through 3 of the OSI reference model (e.g., protocol handling, multiplexing, and switching). These functions are typically executed by equipment such as PBXs, LANs, terminal cluster controllers and multiplexers. Network Termination 1,2 (NT12) is a single piece of equipment that combines the functionality of NT1 and NT2. Terminal Equipment (TE) provides functions undertaken by such terminal equipment as digital telephones, data terminal equipment and integrated voice/data workstations. There are two types of TEs, namely TE1 and TE2. TE1 refers to devices that support standard ISDN interface, while TE2 are those which don't directly support ISDN interfaces. Such non-ISDN interfacing equipment requires Terminal adapters (TA) to connect into ISDN facility.

The reference points define the interface between the functional groups and these include: Rate (R), System (S), Terminal (T), and User (U) reference points. The R reference point is the functional interface between a non-ISDN terminal and the terminal adapter. The S reference point is the functional interface seen by each ISDN terminal. The T reference point is the functional interface seen by the user's of NT1 and NT2 equipment. The U reference point defines the functional interface between the ISDN switch and the network termination equipment (NT2). Standardization of this reference point is essential, especially when NT1s and the Central Office modules are manufactured by different vendors.

It is a generally accepted fact that ISDN can not only be used as a separate entity, but also as tributary network and can play an important role in hybrid networks. So applications that have been traditionally provided by different networking schemes can now be provided in conjunction with ISDN. Some typical applications for Video in the enterprise wide networks include video-telephony in 2B+D circuit switched networks, video conferences over public H0 and H11 links, reconfigure private video conferences networks over channel switched/permanent H0 and H11 links. Medical imaging over 23B+D networks is also one of the many partial lists of ISDN applications.
3.4.2 Broadband Integrated Service Data Network (B-ISDN)

With the explosive growth of network applications and services, it has been recognized that ISDN's limited bandwidth cannot deliver the required bandwidth for these emerging applications. Consequently, the majority of the delegates within CCITT COM XVIII agreed that there is a need for a broadband ISDN (B-ISDN) that allows total integration of broadband services in 1985. And since then, the original ISDN is referred as Narrow-band ISDN (N-ISDN).

The selected transfer mode for B-ISDN has changed several times since its inception. So far two types of transfer modes have been used for digital data transmission: Synchronous Transfer Mode (STM) and Asynchronous Transfer Mode (ATM). STM is suitable for traffic that has severe real time requirements (e.g., voice and video traffic). This mode is based on circuit switching service in which the network bandwidth is divided into periodic slots. Each slot is assigned to a call according to the peak rate of the call. However, this protocol is rigid and does not support bursty traffic. The size of data packets transmitted on a computer network varies dynamically depending on the current activity of the system. Furthermore, some traffic on a data communication network is time insensitive. Therefore, the STM is not selected for B-ISDN and the ATM, a packet switching technique, is selected.

ATM technology divides voice, data, image, and video into short packets, and transmits these packets by interleaving them across an ATM link. The packet transmission time is equal to the slot length. In ATM the slots are allocated on demand, while for STM periodic slots are allocated for every call. In ATM, therefore, no bandwidth is consumed unless information is actually transmitted.

Another important parameter is whether the packet size should be fixed or variable. The main factors that need to be taken into consideration when we compare fixed packet size vs. variable packet sizes are the transmission bandwidth efficiency, the switching performance (i.e. the switching speed, and the switch's complexity) and the delay. Variable packet length is preferred to achieve high transmission efficiency. Because with fixed packet length, a long message has to be divided into several data packets. And each data packet is transmitted with overhead. Consequently, the total transmission efficiency would be low. However, with variable packet length, a long message can be transmitted with only one overhead. Since the speed of switching depends on the functions to be performed, with fixed packet length, the header processing is simplified, and therefore the processing time is reduced. Consequently, from switching point of view, fixed packet length is preferable.

From delay perspective, the packets with fixed small size result in minimal functionalities at intermediate switches and take less time in queue memory management; As a result, fixed size packets reduce the experienced delays in the
overall network. For broadband network, with large bandwidth, the transmission efficiency is not as critical as high-speed throughput and low latency. The gain in the transmission efficiency brought by the variable packet length strategy is traded off for the gain in the speed and the complexity of switching and the low latency brought by the fixed packet length strategy. In 1988, the CCITT decided to use fixed size cells in ATM.

Another important parameter that the CCITT needed to determine, once it decided to adopt fixed size cells, is the length of cells. Two options were debated in the choice of the cell length, 32 bytes and 64 bytes. The choice is mainly influenced by the overall network delay and the transmission efficiency. The overall end-to-end delay has to be limited in voice connections, in order to avoid echo cancellers. For a short cell length like 32 bytes, voice connections can be supported without using echo cancellers. However, for 64 byte cells, echo cancellers need to be installed. From this point of view, Europe was more in favor of 32 bytes so echo cancellers can be eliminated. But longer cell length increases transmission efficiency, which was an important concern to the US and Japan. Finally a compromise of 48 bytes was reached in the CCITT SGXVIII meeting of June 1989 in Geneva.

In summary, ATM network traffic is transmitted in fixed cells with 48 bytes as payload and another 5 bytes as header for routing through the network. The network bandwidth is allocated on demand, i.e., asynchronously. The cells of different types of traffic (voice, video, imaging, data, etc.) are interleaved on a single digital transmission pipe. This allows statistical multiplexing of different types of traffic if burst rate exceeds available bandwidth for a certain traffic type. An ATM network is highly flexible and can support high-speed data transmission as well as real-time voice and video applications.

3.5 ATM

3.5.1 Virtual Connections

Fundamentally ATM is a connection-oriented technology, different from other connection-less LAN technologies. Before data transmission takes place in an ATM network, a connection needs to be established between the two ends using a signaling protocol. Cells then can be routed to their destinations with minimal information required in their headers. The source and destination IP addresses, which are the necessary fields of a data packet in a connection-less network, are not required in an ATM network. The logical connections in ATM are called virtual connections.
Two layers of virtual connections are defined by CCITT: virtual channel (VC) connections (VCC) and virtual path (VP) connections (VPC). One transmission path contains several VPs, as shown in Figure 3.9, and some of them could be permanent or semi-permanent.

Furthermore, each VP contains bundles of VCs. By defining VPC and VCC, a virtual connection is identified by two fields in the header of an ATM cell: Virtual Path Identifier (VPI) and Virtual Channel Identifier (VCI).

The VPI/VCI only have local significance per link in the virtual connection. They are not addresses and are used just for multiplexing and switching packets from different traffic sources. Hence ATM does not have the overhead associated with LANs and other packet switched networks; where packets are forwarded based on the headers and addresses that vary in location and size, depending on the protocol used. Instead an ATM switch only needs to perform a mapping between the VPI/VCI of a cell on the input link and an appropriate VPI/VCI value on the output link.

- **Virtual Channel Connection**: Virtual channel connection is a logical end-to-end connection. It is analogous to a virtual circuit in X.25 connection. It is the concatenation of virtual channel links, which exist between two switching points. A virtual channel has traffic usage parameters associated with it, such as cell loss rate, peak rate, bandwidth, quality of service and so on.

- **Virtual Path Connections**: A virtual path connection is meant to contain bundles of virtual channel connections that are switched together as one unit. The use of virtual paths can simplify network architecture and increase network performance and reliability, since the network deals with fewer aggregated entities.
VPI = 5
VCI = 1, 2, 3

VPI = 10
VCI = 5, 6

VPI = 20
VCI = 5, 6

VPI = 8
VCI = 1, 2, 3

VPI = 8
VCI = 4

VPI = 5
VCI = 2

VPI = 5
VCI = 4

VPI = 8
VCI = 8

VPI = 6
VCI = 10

VCI values are unchanged

Both VPI and VCI values can be changed

Figure 3.10: Switching in ATM

The VPI/VCI fields in an ATM cell can be used to support two types of switching: VP switching and VP/VC switching. In a VP switch, the VPI field is used to route the cells in the ATM switch while the VCI values are not changed as shown in Figure 3.10.

3.5.2 B-ISDN Reference Model

The ATM protocol reference model consists of the higher layers, the ATM layer, the ATM Adaptation Layer (AAL), and the Physical layer. The ATM reference/stack model differs from the OSI (Open System Interconnection) model in its use of planes as shown in Figure 3.11. The portion of the architecture used for user-to-user or end-to-end data transfer is called the User Plane (U-Plane). The Control Plane (C-Plane) performs call connection control. The Management Plane (M-Plane) performs functions related to resources and parameters residing in its protocol entities.

Figure 3.11 Layers of ATM
ATM is connection-oriented, and it uses out-of-band signaling. This is in contrast with the in-band signaling mode of the OSI protocols (X.25) where control packets are inter-mixed with data packets. So during virtual channel connection setup, only the control plane is active. In OSI model, the two planes are merged and are indistinguishable.

**ATM Layers**

The ATM Layers are shown in Figure 3.11. We briefly discuss each of the layers.

**Physical Layer**

The Physical Layer provides the transport of ATM cells between two ATM entities. Based on its functionalities, the Physical Layer is segmented into two sublayers, namely Physical Medium Dependent (PMD) sublayer and the Transmission Convergence (TC) sublayer. This sub-layering separates transmission from physical interfacing, and allows ATM interfaces to be built on a variety of physical interfaces. The PMD sublayer is device dependent. Its typical functions include bit timing and physical medium like connectors. TC sublayer generates and recovers transmission frames. The sending TC sublayer performs the mapping of ATM cells to the transmission system. The receiving TC sublayer receives a bit stream from PMD, extracts the cells and passes them on to the ATM layer. It generates and checks HEC (header error control) field in the ATM header, and it also performs cell rate decoupling through deletion and insertion of idle cells.

**Synchronous Optical Network (SONET)**

The SONET (Synchronous Optical NETwork) also known internationally as Synchronous Digital Hierarchy (SDH), is a physical layer transmission standard of B-ISDN (Broadband Integrated Services Digital Network). SONET is a set of physical layers, originally proposed by Bellcore for specifying standards for optic fiber based transmission line equipment. It defines a set of framing standards, which dictates how bytes are transmitted across the links, together with ways of multiplexing existing line frames (T1, T3 etc.) into SONET. The lowest SONET frame rate called STS-1, defines 8 KHz frames of 9 rows and 90 bytes First 3 bytes are used for Operation, Administration and Management (OAM) purposes and the remaining 83 bytes are used for data. This gives a data rate of 51.84 Mbps. The next highest frame rate standard is STS-3 with 9 frames for OAM and 261 bytes for data, providing a 155.52 Mbps data rate. There are other higher speed SONET standards available: STS-12 - 622.08 Mbps, STS-24 - 1.244 Gbps, STS-48 - 2.488 Gbps and so on (STS-N - Nx51.84 Mbps).

The capabilities of SONET are mapped on to a 4-layer hierarchy-namely Photonic (responsible for conversion between electrical and optical signals and specification for physical layer), Section (functionalities between repeater and multiplexer), Line (functionalities between multiplexers) and Path (function between end-to-end transport).
ATM Cell Format

The ATM Layer performs multiplexing and de-multiplexing of cells from different connections (identified by different VPIs/VCIs) onto a single cell stream. It extracts cell headers from received cells and adds cell headers to the cells being transmitted. Translation of VCI/VPI may be required at ATM switches. Figure 3.13 (a) shows the ATM cell format. Cell header formats for UNI (User-Network Interface) and NNI (Network-Network Interface) are shown in Figures 3.13 (b) and (c), respectively. The functions of the various fields in the ATM cell headers are as follows:

- **Generic Flow Control (GFC):** This is a 4-bit field used only across UNI to control traffic flow across the UNI and alleviate short term overload conditions, particularly when multiple terminals are supported across a single UNI.

- **Virtual Path Identifier (VPI):** This is an 8-bit field across UNI and 12-bits across NNI. For idle cells or cells with no information the VPIs are set to zero, this is also the default value for VPI. The use of non-zero values of VPI across NNI is well understood (for trunking purposes), however the procedures for accomplishing them are under study.
### ATM Cell Format

- **Virtual Circuit Identifier (VCI):** The 16-bit VCI is used to identify the virtual circuit in a UNI or an NNI. The default value for VCI is zero. Typically VPI/VCI values are assigned symmetrically; that is, the same values are reserved for both directions across a link.

- **Payload Type Identifier (PTI):** This is a 3-bit field for identifying the payload type as well as for identifying the control procedures. When bit 4 in the octet is set to 0, it means it is a user cell. For user cells, if bit 3 is set to 0, it means that the cell did not experience any congestion in the relay between two nodes. Bit 2 for user cell is used to indicate the type of user cell. When bit 4 is set to 1, it implies the cell is used for management functions as error indications across the UNI.

- **Cell Loss Priority (CLP):** This field is used to provide guidance to the network in the event of congestion. The CLP bit is set to 1 if a cell can be discarded during congestion. The CLP bit can be set by the user or by the network. An example for the network setting is when the user exceeds the committed bandwidth, and the link is under-utilized.
• **Header Error Check (HEC):** This is an 8-bit Cyclic Redundancy Code (CRC) computed over all fields in the ATM cell header. It is capable of detecting all single bit errors and certain multiple bit errors. It can also be used to correct single bit errors, but is not mandatory.

**ATM Adaptation Layer**

The AAL Layer provides the proper interface between the ATM Layer and the higher layers. It enhances the services provided by the ATM Layer according to the requirements of specific applications: real-time, constant bit rate or variable bit rate. Accordingly, the services provided by AAL Layer can be grouped into four classes. The AAL Layer has five types of protocols to support the four classes of traffic pattern. The corresponding relation between the class of service and the type of AAL protocol are as follows.

- **Type 1:** Supports Class A applications, which require constant bit rate (CBR) services with time relation between source and destination. Error recovery is not supported. Examples include real-time voice messages, video traffic and some current data video systems.

- **Type 2:** Supports Class B applications, which require variable bit rate (VBR) services with time relation between source and destination. Error recovery is also not supported. Examples are teleconferencing and encoded image transmission.

- **Type 3:** Supports Class C applications, which are connection-oriented (CO) data transmission applications. Time relation between source and destination is not required. It is intended to provide services to the applications that use a network service like X.25.

- **Type 4:** Supports Class D applications, which are connection-less (CL) data transmission applications. Time relation between source and destination is not required. The current datagram networking applications like TCP/IP or TP4/CLNP belong to Class D. Since the protocol formats of AAL type 3 and type 4 are similar, they have been merged to AAL type ¾.

- **Type 5:** This type was developed to reduce the overhead related to AAL type ¾. It supports connection-oriented services more efficiently. It is more often referred to as ‘Simple and Efficient AAL’, and it is used for Class C applications.

The AAL layer is further divided into 2 sublayers: the convergence sublayer (CS) and the segmentation-and-reassembly sublayer (SAR). The CS is service dependent and provides the functions needed to support specific applications using AAL. The SAR sublayer is responsible for packing information received
An important character in ATM traffic is its burstiness, meaning that some traffic sources may generate cells at a near-peak rate for a very short period of time and immediately afterwards it may become inactive, generating no cells. Such a bursty traffic source will not require continuous allocation of bandwidth at its peak rate. Since an ATM network supports a large number of such bursty traffic sources, statistical multiplexing can be used to gain bandwidth efficiency, allowing more traffic sources to share the bandwidth. But if a large number of traffic sources become active simultaneously, severe network congestion can result.

In an ATM network, congestion control is performed by monitoring the connection usage. It is called source policing. Every virtual connection (VPC or VCC) is associated with a traffic contract which defines some traffic characteristics such as peak bit rate, mean bit rate, and duration of burst time. The network monitors all connections for possible contract violation. It is also a preventive control strategy. Preventive control does not wait until congestion actually occurs. It tries to prevent the network from reaching an unacceptable level of congestion by controlling traffic flow at entry points to the network.

3.6 Peripheral Area Networks (PAN)

The current advances in computer and networking technology are changing the design of the networks that interconnect computers with their peripherals. The use of distributed computing systems allow users to transparently share and access remote computing and peripheral resources available across the network. Hence the complexity of the Local Peripheral Network (LPN)[Cummings, 1990]. Furthermore, the increased processing power of computers has also lead to a significant increase in input/output bandwidth and in the number of required channels. Applications performing intensive scientific, multimedia or database work demand an increase in the input/output bandwidth of computers and peripherals.

The current input/output peripheral standards cannot meet the required input/output bandwidth. Even the cost of cabling and connections represent a significant portion of the total system cost. The specifications of these standards are as follows:

- **Small Computer Systems Interface (SCSI):** This interface is enabled with two features - 1) a base SCSI designed to support low end systems, with a speed of 8-16 Mbps, and 2) a differential SCSI, designed to support middle systems, connects 8 units, over a distance of 25 meters, with a speed of 32 Mbps.
- **Intelligent Peripheral Interface (IPI):** Designed to support middle systems, IPI connects a channel to eight control units, over a distance of 75 meters, with a speed of 48-80 Mbps.

- **IBM Block Mux (OEMI):** Designed to support high end systems. A channel can connect up to 7 units, over a distance of 125 meters, with a speed of 24-58 Mbps.

- **High Performance Peripheral Interface (HIPPI):** Designed to meet the needs of supercomputing applications, the HIPPI channel can deliver 800 Mbps, over a 32-bit parallel bus, whose length can be up to 25 meters.

- **InfiniBand Architecture (IBA):** IBA defines a System Area Network (SAN) for connecting multiple independent processor platforms (i.e., host processor nodes), I/O platforms, and I/O devices. The IBA SAN is a communications and management infrastructure supporting both I/O and inter-processor communications (IPC) for one or more computer systems. An IBA system can range from a small server with one processor and a few I/O devices to a massively parallel supercomputer installation with hundreds of processors and thousands of I/O devices. Furthermore, the internet protocol (IP) friendly nature of IBA allows bridging to an internet, intranet, or connection to remote computer systems.

The Fiber Channel (FC) is a new standard prepared by the ANSI X3T9.3 committee that aims at providing an efficient LIN network that can operate at speeds of gigabits per second. FC is designed to provide a general transport vehicle supporting all the existing peripheral standards mentioned above. This is achieved through the use of bridges which enable data streams from existing protocols to be supported within the FC sub-network. In this case, the FC provides a replacement of the physical interface later, thereby offering various benefits including improved distance and speed. In this section, we'll focus on the main features of the Fiber Channel and HIPPI standards because of their importance to the development of High Performance Distributed Systems.

### 3.6.1 Fiber Channel Standard

Fiber Channel standard (FCS) has a five-layered structure to reduce the interdependency between the functional areas. This layered approach allows changes in technology to improve the implementation of one layer without affecting the design of other layers. For example, this is clearly illustrated at the FC-1 to FC-0 boundary, where the encapsulated data stream can be transmitted over a choice of multiple physical interfaces and media. The functions performed by each layer are outlined below:
• **FC-4**: Defines the bridges between existing channel protocols (IPI-3, SCSI, HIPPI, Block Mux etc.) and FCS. These bridges provide a continuity of system evaluation and provide a means of protecting the customer's investment in hardware and software while at the same time enabling the use of FCS capabilities.

• **FC-3**: Defines the set of communication services, which is common across all nodes. These services are available to all protocol bridges defined in FC-4 layer.

• **FC-2**: Defines the single frame protocol on which FCS communication is based. It also defines the control and data functions, which are contained within the frame format.

• **FC-1**: Defines the encoding and decoding scheme, which is associated with the transmission frame stream. It specifies the special transmission sequences, which are required to enable communication between the physical interfaces.

• **FC-0**: Defines the physical interface that supports the transmission of data through the FC network. This includes specifications for the fiber, connections, and transceivers. These specifications are based on a variety of media, each designed to meet a range of users, from low to high end implementations.

### 3.6.2 Infiniband Architecture

IBA defines a switched communications fabric allowing many devices to concurrently communicate with high bandwidth and low latency in a protected, remotely managed environment. An endnode can communicate over multiple IBA ports and can utilize multiple paths through the IBA fabric. The multiplicity of IBA ports and paths through the network are exploited for both fault tolerance and increased data transfer bandwidth.

IBA hardware off-loads from the CPU much of the I/O communications operation. This allows multiple concurrent communications without the traditional overhead associated with communicating protocols. The IBA SAN provides its I/O and IPC clients zero processor-copy data transfers, with no kernel involvement, and uses hardware to provide highly reliable, fault tolerant communications.
An IBA System Area Network consists of processor nodes and I/O units connected through an IBA fabric made up of cascaded switches and routers.

IBA handles the data communications for I/O and IPC in a multi-computer environment. It supports the high bandwidth and scalability required for IO. It caters to the extremely low latency and low CPU overhead required for IPC. With IBA, the OS can provide its clients with communication mechanisms that bypass the OS kernel and directly access IBA network communication hardware, enabling efficient message passing operation.

IBA is well suited to the latest computing models and will be a building block for new forms of I/O and cluster communication. IBA allows I/O units to communicate among themselves and with any or all of the processor nodes in a system. Thus an I/O unit has the same communications capability as any processor node.

An IBA network is subdivided into subnets interconnected by routers as illustrated in Figure 3.15. Endnodes may attach to a single subnet or multiple subnets.
An IBA subnet is composed of endnodes, switches, routers, and subnet managers interconnected by links. Each IBT device may attach to a single switch or multiple switches and/or directly with each other.

The semantic interface between the message, data service and the adapter is referred to as IBA verbs. Verbs describe the functions necessary to configure, manage, and operate a host channel adapter. These verbs identify the appropriate parameters that need to be included for each particular function. Verbs are not an API, but provide the framework for the OSV to specify the API.

IBA is architected as a first order network and as such it defines the host behavior (verbs) and defines memory operation such that the channel adapter can be located as close to the memory complex as possible. It provides independent direct access between consenting consumers regardless of whether those consumers are I/O drivers and I/O controllers or software processes communicating on a peer to peer basis. IBA provides both channel semantics (send and receive) and direct memory access with a level of protection that prevents access by non participating consumers.

The foundation of IBA operation is the ability of a consumer to queue up a set of instructions that the hardware executes. This facility is referred to as a work queue. Work queues are always created in pairs, called a Queue Pair (QP), one for send operations and one for receive operations as shown in Figure 3.16. In general, the send work queue holds instructions that cause data to be transferred between the consumer’s memory and another consumer’s memory, and the receive work queue holds instructions about where to place data that is received from another consumer. The other consumer is referred to as a remote consumer even though it might be located on the same node.
The architecture provides a number of IBA transactions that a consumer can use to execute a transaction with a remote consumer. The consumer posts work queue elements (WQE) to the QP and the channel adapter interprets each WQE to perform the operation.

For Send Queue operations, the channel adapter interprets the WQE, creates a request message, segments the message into multiple packets if necessary, adds the appropriate routing headers, and sends the packet out the appropriate port. The port logic transmits the packet over the link where switches and routers relay the packet through the fabric to the destination. When the destination receives a packet, the port logic validates the integrity of the packet. The channel adapter associates the received packet with a particular QP and uses the context of that QP to process the packet and execute the operation. If necessary, the channel adapter creates a response (acknowledgment) message and sends that message back to the originator.

Reception of certain request messages cause the channel adapter to consume a WQE from the receive queue. When it does, a CQE corresponding to the consumed WQE is placed on the appropriate completion queue, which causes a work completion to be issued to the consumer that owns the QP.

The devices in an IBA system are classified as: Switches, Routers, Channel Adapters, Repeaters, and Links that interconnects switches, routers, repeaters, and channel adapters.

The management infrastructure includes subnet managers and general service agents.
IBA provides Queue Pairs (QP). The QP is the virtual interface that the hardware provides to an IBA consumer and it provides a virtual communication port for the consumer. The architecture supports up to $2^{24}$ QPs per channel adapter and the operation on each QP is independent from the others. Each QP provides a high degree of isolation and protection from other QP operations and other consumers. Thus a QP can be considered a private resource assigned to a single consumer.

The consumer creates this virtual communication port by allocating a QP and specifying its class of service. IBA supports the services shown in Table 3.2.

<table>
<thead>
<tr>
<th>Service Type</th>
<th>Connection Oriented</th>
<th>Acknowledged</th>
<th>Transport</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reliable Connection</td>
<td>yes</td>
<td>Yes</td>
<td>IBA</td>
</tr>
<tr>
<td>Unreliable Connection</td>
<td>yes</td>
<td>no</td>
<td>IBA</td>
</tr>
<tr>
<td>Reliable Datagram</td>
<td>no</td>
<td>Yes</td>
<td>IBA</td>
</tr>
<tr>
<td>Unreliable Datagram</td>
<td>no</td>
<td>no</td>
<td>IBA</td>
</tr>
<tr>
<td>RAW Datagram</td>
<td>no</td>
<td>no</td>
<td>Raw</td>
</tr>
</tbody>
</table>

Table 3.2 Service Types supported by QP

Discussion

Gigabit networks represent a change in kind, not just degree. Substantial progress must be made in the areas of protocol, high-speed computer interfaces and networking equipment. The challenge of the 90's will be resolving these problems as well as providing the means for disparate networking approaches to communicate with each other smoothly and efficiently.Ultimately, however, it seems that we are moving towards a single, public, networking environment built on an infrastructure of gigabit-speed fiber optic links, most probably defined at the physical layer by a FDDI or Sonet standard. ATM protocol support multimedia information like voice, video and data in one integrated networking environment. Infiniband supports connections between many nodes and processors spanning many networks and for applications requiring different QoS.

3.7 Wide Area Networks (WANs)

WANs are built to provide communication solutions for organizations or people who need to exchange digital information between two distant places. Since the distance is large, the local telecommunication company is involved, in fact, WANs are usually maintained by the country's public telecommunication companies (PTT's - like AT&T, Sprint), which offer different communication services.

The main purpose of a WAN is to provide reliable, fast and safe communication between two or more places (Nodes) with low delays and at low prices. WANs
enable an organization to have one integral network between all its departments and offices, even if they are not all in the same building or city, providing communication between the organization and the rest of the world. In principle, this task is accomplished by connecting the organization (and all the other organizations) to the network nodes by different types of communication strategies and applications. Since WANs are usually developed by the PTT of each country, their development is influenced by each PTT's own strategies and politics.

The basic WAN service that the PTT usually offers (for many years) is a Leased Line. A Leased Line is a point-to-point connection between two places, implemented by different transmission media (usually through PSTN Trunks), which creates one link between its nodes. An organization whose networks are based on such lines has to connect each office with one line, meaning that each office is connected to as many lines as the number of offices it is connected to.

The Packet Switched WAN appeared in the 1960's, and defined the basis for all communication networks today. The principle in Packet Switched Data Network (PSDN) is that the data between the nodes is transferred in small packets. This principle enables the PSDN to allow one node to be connected to more than one other node through one physical connection. That way, a fully connected network, between several nodes, can be obtained by connecting each node to one physical link. Another advantage for Packet Switching was the efficient use of resources by sharing the Network bandwidth among the users (instead of dividing).

The first communication Packet Switched Networks were based on the X.25 packet switching protocol. X.25 networks became the de facto standard for nonpermanent data communication and was adopted by most PTT’s.

X.25 networks enabled cheaper communication, since their tariff was based on the communication time and the amount of data transferred. X.25 networks used the PTT's transmission networks more efficiently since the bandwidth was released at the end of the connection, or when no data was transmitted. Another advantage of X.25 was that it allowed easy implementation of international connections enabling organizations to be connected to data centers and services throughout the world. By the 1980's, X.25 networks were the main international channel for commercial data communication.

Today to meet the high speed demands, the WANs rely on technologies ATM (B-ISDN) Frame Relay, SONET and SDH. We have already discussed ATM, SONET and SDH in the previous section on MAN.

3.8 Wireless LANS

3.8.1 Introduction
Wireless LANs provide many convenient facilities that are not provided by traditional LANs such as mobility, relocation, ad hoc networking and coverage of locations difficult to wire. Wireless LANs were not of much practical use until the recent past, due to many technological and economical reasons such as high prices, low bandwidth, transmission power requirements, infrastructure and licensing. These concerns have been adequately addressed over the last few years, and the popularity of wireless LANs are increasing rapidly, and a new standard, namely IEEE 802.11, attempts to standardize these efforts.

### 3.8.2 IEEE 802.11

IEEE 802.11 defines a number of services that wireless LANs are required to provide functionality equivalent to wired LANs. The services specified in this standard are:

1. **Association:** Establishes an initial association between a station and an access point. A LAN's identity and address must be known and confirmed before the station can start transmitting.

2. **Re-association:** Enables an established connection to be transferred from one access point to another, allowing a mobile user to move from one station to another.

3. **Disassociation:** A notification from either a station or an access point that an existing association is terminated. A mobile station has to notify the base station before it shuts down; however, the base stations have a capability to protect themselves against stations that shut down without any notification.

4. **Authentication:** Used to establish the identity of the stations to each other. In a wired LAN, the physical connection always conveys the identity of other station. Here, an authentication scheme has to be used to establish the proper identity of the stations. Though the standards do not specify any particular authentication scheme, the methods used can range from the relatively unsecure handshaking to a public-key encryption scheme.

5. **Privacy:** used to prevent the broadcasted messages being read by users other than the intended recipient. The standard provides for an optional use of encryption to provide a high level of privacy.

The 802.11 standard also specifies three kinds of physical media standards for wireless LANs:

- Infrared at 1 and 2 Mbps operating at a wavelength between 850 and 950nm
- Direct-sequence spread-spectrum operating in the 2.4-GHz ISM band, up to seven channels with data rates of 1 or 2 Mbps can be used.

- Frequency hopping spread spectrum operating in the 2.4-GHz ISM band

### 3.8.3 Classification

Wireless LANs are classified according to the transmission techniques used, and all current wireless LAN products fall into one of the three following categories:

1. **Infrared LANs:** In this case, an individual cell of an IR LAN is limited to a very small distance, such as a single room. This is because infrared rays cannot penetrate opaque walls.

2. **Spread Spectrum LANs:** This makes use of spread-spectrum technology for transmission. Most of the networks in this category operate in the bands where no licensing is required from the FCC.

3. **Narrowband Microwave:** These LANs operate at very high microwave frequencies, and FCC licensing is required.

### 3.8.4 Applications of Wireless LAN Technology

1. **Nomadic Access**

Nomadic access refers to the wireless link between a wireless station and a mobile terminal (a notebook or laptop) equipped with an antenna. This gives access to users who are on the move, and wish to access the main hub from different locations.

2. **Ad-Hoc Networking**

An ad-hoc network is a network set-up temporarily to meet some immediate need, such as conferences and demonstrations. No infrastructure is required for an ad-hoc network, and with the help of wireless technologies, a collection of users within range of each other may dynamically configure themselves into a temporary network.

3. **LAN Extension**

A wireless LAN saves the cost of installation of LAN cabling and eases the task of relocation and other modifications and extensions to the existing network structure.
4. Cross Building Interconnect

Point-to-point wireless links can be used to connect LANs in nearby buildings, independent of whether the buildings within themselves have wired or wireless LANs. Though this is not a typical use of wireless LANs, this is also usually included as an application for the sake of completeness.

Summary

Computer networks play an important role in the design of high performance distributed systems. We have classified computer networks into four types: Peripheral Area Networks, Local Area Networks, Metropolitan Area Networks and Wide Area Networks. For each class, we discussed the main computer technology that can be used to build high performance distributed systems. These networks include HIPPI, Fiber Channel, FDDI, DQDB, and ATM. Gigabit networks represents a change in kind, not just degree. Substantial progress must be made in the areas of protocols, high speed computer interfaces and networking equipment before they can be widely used. The challenge of the 90's will be resolving these problems as well as providing the means for disparate networking approaches to communicate with each other smoothly and efficiently. Ultimately, however, it seems that we are moving toward a single, public networking environment built on an infrastructure of gigabit-speed fiber optic links, most probably defined at the physical layer by a merged Fiber Channel/Sonet standard. ATM protocols will become popular format for sharing multimedia voice, video, and data in one integrated networking environment. Further, ATM has the potential to implement all classes of computer networks and provide the required current and future communication services.

Problems

1. In some cases FDDI synchronous transmission causes a "$\text{\`glitch}"$ in an audio or video signal that must be transmitted at a periodical rate. Describe a scenario where this glitch can occur and suggest a solution to solve this problem.

2. If source computer A wants to send a file of size 10 Kbytes to destination computer B and communication has to take place over a HIPPI channel. Explain how this exchange would take place, with regard to data framing and physical layer control signal sequencing.

3. Discuss the following: Enumerate the appropriate place for FDDI protocol into the OSI reference model. The use of claim frames in the ring initialization process. The
guarantee of TTRT (target token rotation time), used by the FDDI protocol, for maximum delay on the ring. Enumerate

4. How can the Target Token Rotation Time affect the performance of a FDDI network?

5. What are the advantages of using copper as the transmission medium for FDDI?

6. What are the differences between FDDI II and FDDI? How is the former an improvement over the latter?

7. What is the most cost efficient configuration for a large FDDI network? Why?

8. What are the differences between Ethernet, FDDI and Token Ring standards for forming Local Area Networks?

9. DQDB MAC protocol is biased towards the nodes that are close of the slot generators. Explain this scenario and describe one technique to make DQDB protocol more fair.

10. ATM mainly uses virtual connections for establishing a communication path between nodes. Describe the relationship between virtual connection, virtual path, and virtual channel. Based on this relationship, how are the VPI and VCI identifiers used in performing cell switching in the ATM switches. Do you think the sizes of VPI and VCI fields are large enough for holding the required switching information?

11. What are the advantages of having a large packet size? What are the advantages of having a small packet size? Why does ATM prefer smaller packet sizes?

12. What are the different classes supported by the ATM Adaptation Layer? On what basis are the classes divided into, and how?

13. ISDN offers three main services. Describe these services and their applications to real life examples.

14. What is the SONET? Describe its capabilities and limitations.

15. What are the main characteristics of HIPPI. Discuss functions of HIPPI-FP (framing protocol) and HIPPI-LE (link encapsulation).

References


2. Stallings, W., `Local & Metropolitan Area Networks', Macmillan, 1995


5. CCITT COM XVIII-R1-E, February 1985


25. Fiber Channel Standard-XT/91-062.


42. Infiniband Architecture Specification, Volume 1, Release 1.1, Infiniband Trade Association


46. Gopal Agrawal, Biao Chen, and Wei Zhao. Local synchronous capacity allocation schemes for guaranteeing message deadlines with the timed token protocol. In IEEE Infocom 93, volume 1, 1993.