CONTENTS

CHAPTER

A Com
be a comp
other node
commun

1.1 NE

1. Fo
2. Fo
3. Fo
4. Fo
5. Sc
ter
6. Re
7. Ge
loc
8. Edu

2 CLA
GEC

• LA
tens
• MA
kilo
• WA

3 COM

ere are n

CHAPTER 1

Review of Physical Layer and Data Link Layer

A Computer network is a set of nodes connected by means of communication links. A node can be a computer, printer, or any other device capable of sending and/or receiving data generated by other nodes on the network. The links can be of any form like twisted pair, coaxial cable,**microwaves, a communication satellite, etc.**

1.1 NEED FOR COMPUTER NETWORKS

1. For sharing of hardware resources like printers, scanners, plotters, modems, etc.
2. For sharing of software resources like OS, compilers, databases etc.
3. For exchange data and information across domains.
4. For sharing of documents, files, pictures, videos, etc.
5. Scalability: The network can be augmented and upgraded, as the requirement grows. More terminals and hardware resources can be added later on.
6. Reduced Costs: Networks help in reducing the costs by resource sharing and scalability.
7. Geographical Interconnection: Computer Networks allow machines in different geographical locations to share and exchange information.
8. Educational Purposes: Networks are used in school and college for educational purposes.

.2 CLASSIFICATION OF COMPUTER NETWORKS ON THE BASIS OF GEOGRAPHICAL AREA

- LAN : A local area network typically interconnects hosts that are up to a few or maybe a few tens of kilometres apart.
- MAN : A metropolitan area network typically interconnects devices that are up to a few hundred kilometres apart
- WAN : A wide area network interconnects hosts that can be located anywhere on Earth

.3 COMPUTER NETWORK APPLICATIONS

here are numerous applications of computer networks. Today we cannot even imagine a world

without these networks. These applications are very diverse and range from simple file sharing to parallel and distributed applications such as banking. Some of these applications include:

1. Instant Messaging
2. Remote Login
3. Email
4. File sharing
5. Sharing video clips
6. Online Shopping
7. Web and Information Sharing
8. Providing services like IT Return, property tax, passport service, driving licence, etc.
9. Video-conferencing
10. Voice over Internet Protocol (VoIP)
11. B2B applications like order entry, centralized purchasing, inventory control, etc
12. B2C applications like Airline and Train Reservation, Hotel booking and car rental
13. Banking Applications
14. Stock Market

1.4 NETWORK TOPOLOGY

The topology of a network describes the physical connection structure between the nodes of communication network. This kind of connection structure determines the implementation of t physical network, the limits of applications, and the parameters of the networks.

Types of Network Topology

Star Topology

All are connected to a central node via point-to-point connections.

Advantage	Disadvantage
Directly connected to central node from every node	Generally high tota3l length of connections if node are ordered as geographical line
Simple integration for more nodes	Central node requires N interfaces for N nodes
Easy to implement with optical transmission media	Communication between nodes ONLY possible through central node
	If central node fails, no communication possible

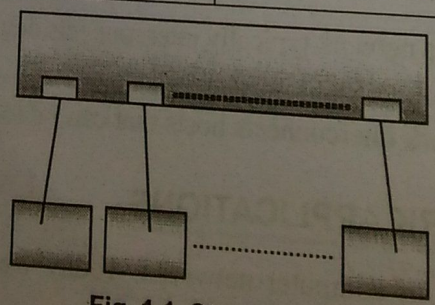

Fig. 1.1. Star topology

Bus Topology

It is commonly

Lower cabling cost
Easy connecting to
Simple to extend n
Failure of one node
Arbitrary logical c

ree Topology

When arbitrary

Low cabling

ng Topology

ing topology is

ssible impleme
ry well suited f edia
mple node iden

Bus Topology

It is commonly known by the electrically passive connection of all nodes to a common medium.

Advantage	Disadvantage
Lower cabling costs	Limited bus length and number of nodes
Easy connecting to a node	Stub length for connecting to nodes could be strictly limited due to terminations of resistors
Simple to extend more nodes without interruption	Complicated to implement with optical media
Failure of one node does not affect the rest Arbitrary logical communication possible	Node identification required

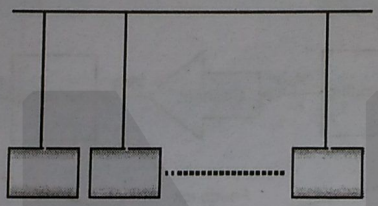

Fig. 1.2. Bus topology

Tree Topology

When arbitrary branching is possible via active or passive elements, it is called tree topology.

Advantage	Disadvantage
Low cabling and installation costs	When active branching element are used, their cost are a disadvantage
	Possible to have branches of considerable length

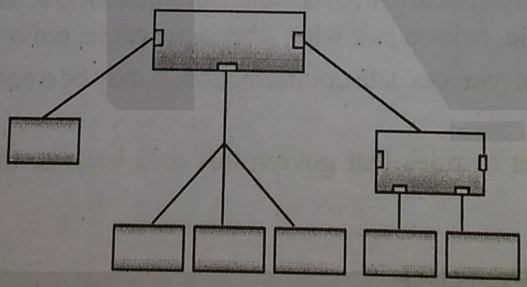

Fig. 1.3. Tree topology

Ring Topology

A ring topology is defined by a closed chain of addressed point-to-point connections.

Advantage	Disadvantage
Possible implementation of extended network	Total system fails when one of the node fails
Very well suited for the use of optical transmission media	When integrate a new node or replace a node, interruption is needed
Simple node identification possible of nodes possible	

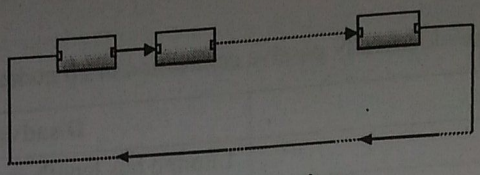

Fig. 1.4. Ring topology

1.5 THE COMMUNICATION PROCESS

Data Communications: It is defined as the exchange of data between two devices using some transmission media.

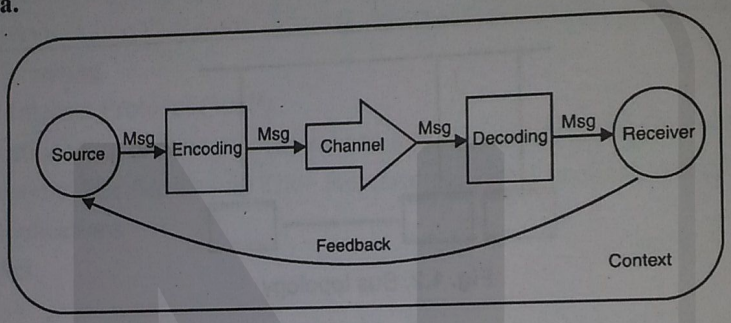

Fig. 1.5. The Communication Process

The main components of a data communication system are:

1. **Message/Data: It is the data which we want to transmit** from one point to another. It may b a sound, text, number, movies, videos etc.
2. **Sender** is a device which sends the message. It may be a computer, workstation, video camer mobile etc.
3. **Medium/ Channel** *i.e.* a physical path over which information is send from sender to the receive It may be a coaxial cable, twisted pair wire, fiber optic cable, radio waves used by mobiles et
4. **Receiver** is a device that receives data or information. It may be a computer, mobile, workstatio etc.
5. **Protocol** refers to a set of rules that govern this data transfer through the Communicatio medium.

1.6 LAYERED ARCHITECTURE

Most of the networks are organized in the form of layers. This layered architecture simplifies th network design and reduces the complexity. Some important properties of the layered architecture ar

1. The number of layers, their nomenclature and the function of each layer varies from one netwo to another.
2. Each layer offers some services to the next higher layers.
3. A Layer X on one machine (source) carries on a conversation with layer X only on the destinatio machine.
4. The rules and conventions used in this conversation are collectively known as the layer protocol.

5. A protocol is an agreement between the two machines as how communication link should be established, maintained and released.
6. The communication protocols are necessary for communication, otherwise communication it is not possible.

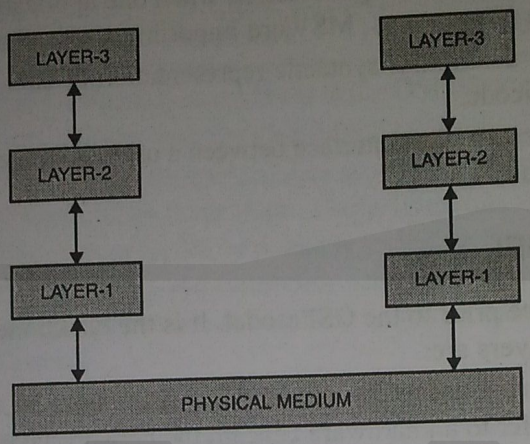

Fig. 1.6. Two main network Layer Architectures

There are two main Layered Architectures:

1. The seven-layer OSI/ISO model – Open Systems Interconnection, currently maintained by the International Organization for Standards. It was first introduced in 1970's. It has a total of 7 Layers.
2. The four-layer TCP/IP model – Transmission Control Protocol/Internet Protocol.

The relative comparison of the two models is shown in figure:

1.7 OSI REFERENCE MODEL

There are 7 layers in this model, where each layer offers some services to higher layers. These services are provided through the interfaces between these layers.

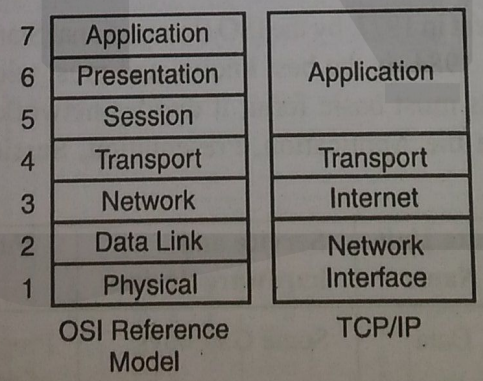

Fig. 1.7. Relative comparison of OSI and TCP/IP Model

1. Physical layer – controls electrical and mechanical aspects of data transmission, e.g., voltage levels, cable lengths, and so on.
2. Data-link layer – addresses the transmission of data frames (or packets) over a physical link between network entities, includes error correction.

3. Network layer – establishes paths for data between computers and determines switching among routes between computers, determines how to disaggregate messages into individual packets
4. Transport layer – deals with data transfer between end systems and determines flow control
5. Session layer – creates and manages sessions when one application process requests access to another applications process, e.g., MSWord importing a spread sheet from Excel.
6. Presentation layer – determines syntactic representation of data, e.g., agreement on character code like ASCII/Unicode.
7. Application layer – establishes interface between a user and a host computer, e.g., searching a database application.

1.8 THE TCP/IP MODEL

This model is in existence prior to the OSI model. It is the backbone of the Internet. The TCP/IP model has 4 layers. These layers are:

1. Host to Network layer – not really part of this model, since TCP and IP deal with software usually thought to refer to all hardware beneath the network layer. It supports all standard and proprietary protocols of physical and Data-Link layer.
2. Internet or network layer – provides network addressing and routing, providing a common address space and connecting heterogeneous networks. IP runs here.
3. Transport layer – manages data-consistency by providing a reliable [two meanings!!] byte stream between nodes on a network. TCP and User Datagram Protocol (UDP) run here.
4. Process and applications layer – provides application services to users and programs.

1.9 NETWORKING IN THE OSI MODEL

It is very important to learn the OSI model, since it is the foundation to understanding the network world. The Open Systems Interconnection Reference Model (OSI Reference Model or OSI Model) an abstract description for layered communications and computer network protocol design. It is used to describe the flow of data between the physical connection to the network and the end-user application.

This model initial development in 1977, by the ISO (International Standards Organization), redesigned and released for general use in 1984, is the best known and most widely used model for describing networking environments. In its most basic form, it divides network architecture into seven layers which, from top to bottom, are the Application, Presentation, Session, Transport, Network, Data Link, and Physical Layers.

Layer Number	Layer Name	Data Unit Name	Service and hardware devices	Functionality
7	Application	Data	Some Gateways	Program to program transfer of data
6	Presentation	Data	Redirector Service	Displaying the information
5	Session	Data		Establishing, maintaining and coordinating communication
4	Transport	Segment		Accurate delivery, service quality

3	Network	Packet	Most Gateways, Routers	Transport routes, message handling and transfer
2	Data Link	Frame	Switches, bridges	Coding, addressing, and transmitting information
1	Physical	Bit	Repeaters	Hardware connections

Layer 7 - Application Layer: The application layer is the OSI layer closest to the end user, which means that both the OSI application layer and the user interact directly with the software application.

Layer 6 - Presentation Layer: The presentation layer provides a variety of coding and conversion functions that are applied to application layer data. These functions ensure that information sent from the application layer of one system would be readable by the application layer of another system. Some examples of presentation layer coding and conversion schemes include common data representation formats, conversion of character representation formats, common data compression schemes, and common data encryption schemes.

Layer 5 – Session Layer: The session layer establishes, manages, and terminates communication sessions. Communication sessions consist of service requests and service responses that occur between applications located in different network devices.

Layer 4 - Transport Layer: The transport layer accepts data from the session layer and segments the data for transport across the network. Generally, the transport layer is responsible for making sure that the data is delivered error-free and in the proper sequence. Flow control generally occurs at the transport layer.

Layer 3 - Network Layer: The network layer defines the network address, which differs from the MAC address. Some network layer implementations, such as the Internet Protocol (IP), define network addresses in a way that route selection can be determined systematically by comparing the source network address with the destination network address and applying the subnet mask.

Layer 2 – Data Link Layer: The data link layer provides reliable transit of data across a physical network link. Different data link layer specifications define different network and protocol characteristics, including physical addressing, network topology, error notification, sequencing of frames, and flow control. Network topology consists of the data link layer specifications that often define how devices are to be physically connected, such as in a bus or a ring topology.

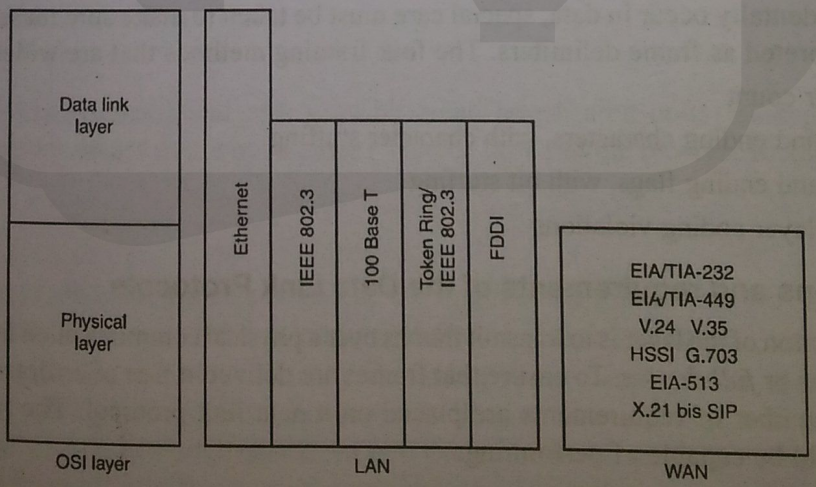

Fig. 1.8. Physical layer implementations

Layer 1 – Physical Layer: The physical layer defines the electrical, mechanical, procedural, a functional specifications for activating, maintaining, and deactivating the physical link betwe communicating network systems. Physical layer specifications define characteristics such as volta levels, timing of voltage changes, physical data rates, maximum transmission distances, and physic connectors. Physical layer implementations can be categorized as either LAN or WAN specificatio The following figure illustrates some common LAN and WAN physical layer implementations.

1.10 DATA LINK LAYER

What is DLL (Data Link Layer)

The Data Link Layer is the second layer in the OSI model, above the Physical Layer, whi ensures that the error free data is transferred between the adjacent nodes in the network. It breaks t datagrams passed down by above layers and convert them into frames ready for transfer. This called Framing. It provides two main functionalities

- Reliable data transfer service between two peer network layers
- Flow Control mechanism which regulates the flow of frames such that data congestion is there at slow receivers due to fast senders.

1.10.1 Purpose of the Data Link Layer

- Data Link layer provides structure to the 1s and 0s that are transmitted over the transmiss media in the physical layer thus providing meaningful data between the upper layers of sending and receiving nodes. The TCP/IP network access layer is the equivalent of the Data link (Layer 2) and Physical (Layer 1).
- The data link layer is responsible for the exchange of frames between the network devices o a physical network. It enables the upper layers to access the transmission media and contr how data is transmitted and received on the network.

1.10.2 What is Framing?

Since the physical layer merely accepts and transmits a stream of bits without any regard meaning or structure, it is upto the data link layer to create and recognize frame boundaries. This be accomplished by attaching special bit patterns to the beginning and end of the frame. If these patterns can accidentally occur in data, special care must be taken to make sure these patterns are incorrectly interpreted as frame delimiters. The four framing methods that are widely used are

- Character count
- Starting and ending characters, with character stuffing
- Starting and ending flags, with bit stuffing
- Physical layer coding violations

1.10.3 Functions and requirements of the Data Link Protocols

The basic function of the layer is to transmit frames over a physical communication link. Transmis may be *half duplex* or *full duplex*. To ensure that frames are delivered free of errors to the destina station (IMP) a number of requirements are placed on a data link protocol. The protocol (co mechanism) should be capable of performing:

1. The identification of a frame (i.e. recognise the first and last bits of a frame).
2. The transmission of frames of any length up to a given maximum. Any bit pattern is permitted in a frame.
3. The detection of transmission errors.
4. The retransmission of frames which were damaged by errors.
5. The assurance that no frames were lost.
6. In a multidrop configuration Some mechanism must be used for preventing conflicts caused by simultaneous transmission by many stations.
7. The detection of failure or abnormal situations for control and monitoring purposes.

It should be noted that as far as layer 2 is concerned a host message is pure data, every single bit of which is to be delivered to the other host. The frame header pertains to layer 2 and is never given to the host.

1.10.4 Data Link Layer Frame Format

Data Link Layer PDU is the Frame. Its format is given below: *f*

Frame Start	Addressing	Type	DATA	Error Detection	Frame Stop

NOTE: *Each frame contains exactly one network layer datagram. The maximum size of the Frame data varies as per the data link and physical layer protocols and can be measured in Maximum Transmission Unit (MTU). e.g. the MTU of Ethernet standard 10/100 Mbps links is 1500 bytes.*

Typical field types include:

a. Start and stop indicator fields - The beginning and end limits of the frame.
b. Naming or addressing fields
c. Type field - The type of PDU contained in the frame
d. Quality control fields
e. A data field -The frame payload (Network layer packet) *f*
f. Frame Check Sequence (FCS) – It checks for bit-errors during transmission

1.10.5. Data Link Layer Functions

Framing

The DLL translates the physical layer's raw bit stream into discrete units (messages) called *frames*. How can the receiver detect frame boundaries? That is, how can the receiver recognize the start and end of a frame?

Length Count

Make the first field in the frame's header be the length of the frame. That way the receiver knows how big the current frame is and can determine where the next frame ends.

Disadvantage: Receiver loses synchronization when bits become garbled. If the bits in the count become corrupted during transmission, the receiver will think that the frame contains fewer (or more) bits than it actually does. Although checksum will detect the incorrect frames, the receiver will have difficulty re-synchronizing to the start of a new frame. This technique is not used anymore, since better techniques are available.

Bit Stuffing

This technique replaces within the frame every occurrence of two consecutive 1's with 110. E.g. append a zero bit after each pair of 1's in the data. This prevents 3 consecutive 1's from ever appearing in the frame.

Similarly, the receiver converts two consecutive 1's followed by a 0 into two 1's, but recognizes the 0111 sequence as the end of the frame.

Example: The frame "1011101" would be transmitted over the physical layer as "01111011010101111".

Note: In Bit stuffing technique locating the start/end of a frame is easy, even when frames are damaged. The receiver simply scans arriving data for the reserved patterns. Moreover, the receiver will resynchronize quickly with the sender as to where frames begin and end, even when bits in the frame get garbled. The main disadvantage with bit stuffing is the insertion of additional bits into the data stream, wasting bandwidth.

Character stuffing

This is similar to bit-stuffing, but operates on bytes instead of bits. It uses reserved characters to indicate the start and end of a frame. For instance, use the two-character sequence DLE STX (Data-Link Escape, Start of TeXt) to signal the beginning of a frame, and the sequence DLE ETX (End of TeXt) to flag the frame's end.

Note: If two-character sequence DLE ETX happens to appear in the frame itself then the following given solution is implemented: We can make use of *character stuffing*; within the frame which replaces every occurrence of DLE with the two-character sequence DLE DLE. The receiver reverses the processes replacing every occurrence of DLE DLE with a single DLE. The disadvantage of this scheme is that it uses character as the smallest unit that can be operated on however not all architectures are byte oriented.

Example: If the frame contained "A B DLE D E DLE", the characters transmitted over the channel would be "DLE STX A B DLE DLE D E DLE DLE DLE ETX".

Encoding Violations

We can send a signal that doesn't conform to any legal bit representation protocol; for example in Manchester encoding, 1-bits are represented by a high-low sequence, and 0-bits by low-high sequence. The start/end of a frame could be represented by the signal low-low or high-high.

The advantage of encoding violations is that no extra bandwidth is required as in bit-stuffing. The IEEE 802.4 standard uses this approach.

Flow Control

Flow control deals with controlling the speed of the sender to match that of the receiver. Usually this is a dynamic process, as the receiving speed depends on various factors such as the load, and availability of buffer space. It may be implemented through various protocols such as:

- Stop and Wait
- Sliding Window Protocol

Link Management

In several cases, the data link layer service must be "opened" before use:

- The data link layer uses open operations for allocating buffer space, control blocks, agreeing on the maximum message size, etc.

- Synchronize and initialize send and receive sequence numbers with its peer at the other end of the communications channel.

Error Control

Error control is concerned with insuring that all frames are eventually delivered (possibly in order) to a destination. It requires three things which are as follows:.

Acknowledgements:

Typically, reliable delivery is achieved using the "acknowledgments with retransmission" paradigm, whereby the receiver returns a ***acknowledgment*** (ACK) frame to the sender indicating the correct receipt of a frame.

In some systems, the receiver also returns *negative acknowledgment* (NACK) for incorrectly-received frames. This acts like a signal for the sender so that it can retransmit a frame right away without waiting for a timer to expire.

Timers:

Simple ACK/NACK schemes fail to address a problem arising due to a frame that is lost, and as a result, fails to solicit an ACK or NACK. This problem is solved by the use of *Retransmission timers* which are used to resend frames that don't produce an ACK. When sending a frame, schedule a timer to expire at some time after the ACK should have been returned. If the timer goes off, retransmit the frame.

Sequence Numbers:

Retransmissions introduce the possibility of duplicate frames. To remove duplicate frames, add sequence numbers to each frame, so that a receiver can distinguish between new frames and old copies.

Error Control Protocols

Protocols listed under this may be summarized as:

- Stop-and-wait ARQ
- Go-Back-N ARQ
- Selective Repeat ARQ

Error Detection and Correction

Detecting and correcting errors requires sending of additional information along with the data. There are two types of techniques which may be used against errors:

Error Detecting Codes:

This requires addition of redundancy bits to *detect* errors and use ACKs and retransmissions to recover from the errors. These may be categorized as follows:

- Parity checks
- Checksum Method
- Cyclic redundancy codes: CRC Checksums: The most popular error detection codes are based on *polynomial codes* or *cyclic redundancy codes*.

CRC Standards

There are currently 3 international standards for CRC:

- CRC-12: $x^{12} + x^{11} + x^3 + x^2 + x + 1$
- CRC-16: $x^{16} + x^{15} + x^2 + 1$
- CRC-CCITT: $x^{16} + x^{12} + x^5 + 1$

The 16-bit CRC detects all single and double errors, all errors with odd number of bits, all burst errors of length £ 16 bits, and 99.997% of 17-bit errors. This is usually implemented in hardware level.

Error Correcting Codes:

This requires addition of redundancy bits to detect *and* correct errors. Eg. ***Hamming code***

1.10.6 Data Link Sublayers

The data link layer is actually divided into two sub-layers:

- **Logical Link Control (LLC):** Also called IEEE 802.2 Protocol, this upper sub-layer decides the software processes that provide services to the network layer protocols. It places information in the frame as per the protocols being used.
- **Media Access Control (MAC):** This is the lower sub-layer of DLL and defines the transmission media access processes performed by the hardware. It performs data link layer addressing and delimiting of data as per the physical layer signalling requirements.

The advantage of dividing the data link layer into sub-layers is that this allows for one type of frame defined by the upper layer to access different types of transmission media defined by the lower layer. For example Ethernet protocol uses this feature.

The following figure depicts division of the data link layer into the LLC and MAC sub-layers. The LLC is responsible for communicating with the network layer while the MAC sub-layer supports different network access technologies. For example, the MAC sub-layer communicates with Ethernet LAN technology to send and receive frames over copper or fiber-optic cable. The MAC sub-layer also communicates with wireless technologies such as Wi-Fi and Bluetooth to send and receive frames wirelessly.

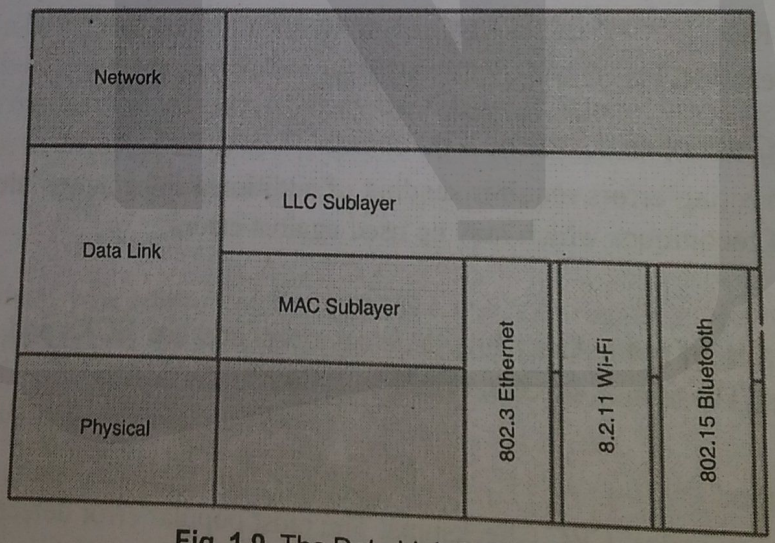

Fig. 1.9. The Data Link Sublayers

1.11 MEDIA ACCESS CONTROL

This Layer 2 protocols specifies the method of encapsulating a packet into a frame and gives details of techniques for getting the encapsulated packet on and off each transmission medium. The technique used for getting the frame on and off the transmission media without data collision is called the media access control method.

When the data packets travel from source node to destination node, they generally travel over different physical networks. These physical networks may be constructed from different types of physical media such as twisted pair cables, copper wires, fiber optic cables, and wireless medium consisting of Bluetooth frequency, electromagnetic signals, radio and microwave frequencies, and satellite links. The media access control methods described by the data link layer protocols define the procedures by which nodes can access the transmission media and transmit frames in diverse network structures.

In the below example, a PC in Paris is transmitting data to a laptop in Japan. Note that the two hosts are communicating using IP exclusively; here the situation might want for the use of numerous data link layer protocols for transmitting the IP packets over various types of LANs and WANs using different protocols. Each transition at a router may require a different data link layer protocol for transport on a new medium.

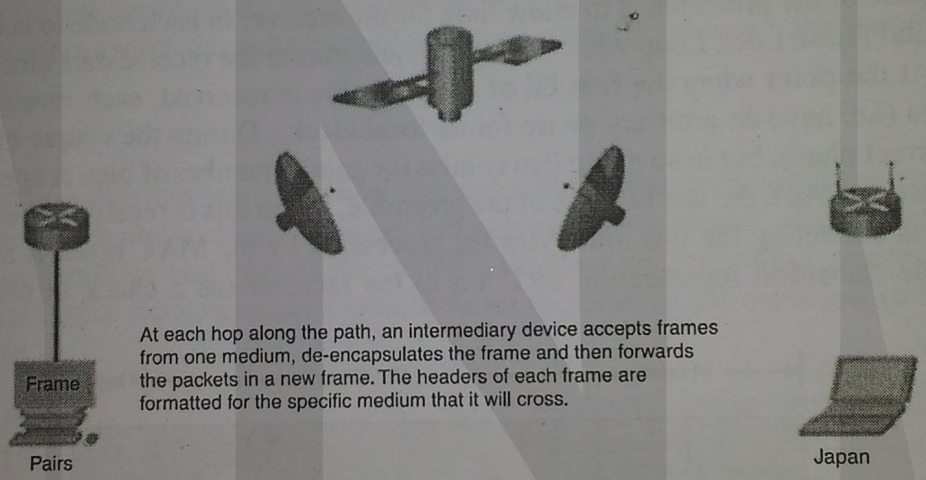

Fig. 1.10. The MAC Layer handles Media Access issues of different Transmission Media

The IEEE 802 suite contains three Medium access protocols.

1. IEEE 802.3 "Ethernet", supported originally by Xerox and DEC.
2. IEEE 802.4 Token Bus. Supported by GM.
3. IEEE 802.5 Token ring. Supported by IBM.

1.11.1 MAC Header

The Medium Access Control (MAC) protocol is used to provide the data link layer of the Ethernet LAN system. The MAC protocol encapsulates a SDU (payload data) by adding a 14 byte header (Protocol Control Information (PCI)) before the data and appending an integrity checksum, The checksum is a 4-byte (32-bit) Cyclic Redundancy Check (CRC) after the data. The entire frame is preceded by a small idle period (the minimum inter-frame gap, 9.6 microsecond (μS)) and a 8 byte preamble (including the start of frame delimiter).

Preamble

The purpose of the idle time before transmission starts is to allow a small time interval for the receiver electronics in each of the nodes to settle after completion of the previous frame. A node starts transmission by sending an 8 byte (64 bit) preamble sequence. This consists of 62 alternating 1's and 0's followed by the pattern 11. Strictly speaking the last byte which finished with the '11' is known as the "Start of Frame Delimiter". When encoded using Manchester encoding, at 10 Mbps, the 62 alternating bits produce a 10 MHz square wave (one complete cycle each bit period).

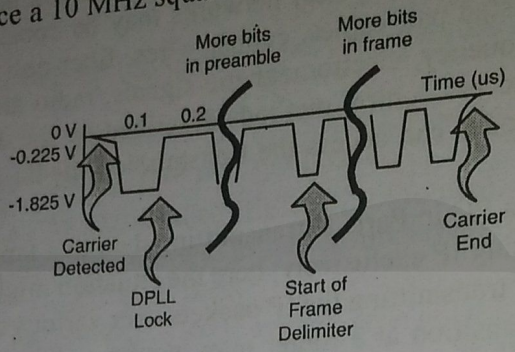

Fig. 1.11.

The purpose of the preamble is to allow time for the receiver in each node to achieve lock of the receiver Digital Phase Lock Loop which is used to synchronise the receive data clock to the transmit data clock. At the point when the first bit of the preamble is received, each receiver may be in an arbitrary state (i.e. have an arbitrary phase for its local clock). During the course of the preamble it learns the correct phase, but in so doing it may miss (or gain) a number of bits. A special pattern (11) is therefore used to mark the last two bits of the preamble. When this is received, the Ethernet receive interface starts collecting the bits into bytes for processing by the MAC layer. It also confirms the polarity of the transition representing a '1' bit to the receiver (as a check in case this has been inverted).

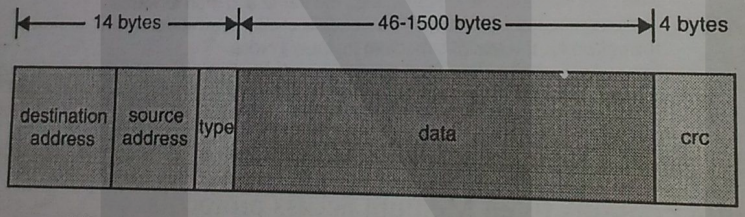

Fig. 1.12. MAC encapsulation of a packet of data

The header consists of three parts:

- A 6-byte destination address, which specifies either a single recipient node (unicast mode), a group of recipient nodes (multicast mode), or the set of all recipient nodes (broadcast mode).
- A 6-byte source address, which is set to the sender's globally unique node address. This may be used by the network layer protocol to identify the sender, but usually other mechanisms are used (e.g. arp). Its main function is to allow address learning which may be used to configure the filter tables in a bridge.
- A 2-byte type field, which provides a Service Access Point (SAP) to identify the type of protocol being carried (e.g. the values 0x0800 is used to identify the IP network protocol, other values are used to indicate other network layer protocols). In the case of IEEE 802.3 LLC, this may

also be used to indicate the length of the data part. Th type field is also be used to indicate when a Tag field is added to a frame.

CRC

The final field in an Ethernet MAC frame is called a Cyclic Redundancy Check (sometimes also known as a Frame Check Sequence). A 32-bit CRC provides error detection in the case where line errors (or transmission collisions in Ethernet) result in corruption of the MAC frame. Any frame with an invalid CRC is discarded by the MAC receiver without further processing. The MAC protocol does not provide any indication that a frame has been discarded due to an invalid CRC.

The link layer CRC therefore protects the frame from corruption while being transmitted over the physical mediuym (cable). A new CRC is added if the packet is forwarded by the router on another Ethernet link. While the packet is being processed by the router the packet data is not protected by the CRC. Router processing errors must be detected by network or transport-layer checksums.

Inter Frame Gap

After transmission of each frame, a transmitter must wait for a period of 9.6 microseconds (at 10 Mbps) to allow the signal to propagate through the receiver electronics at the destination. This period of time is known as the Inter-Frame Gap (IFG). While every transmitter must wait for this time between sending frames, receivers do not necessarily see a "silent" period of 9.6 microseconds. The way in which repeaters operate is such that they may reduce the IFG between the frames which they regenerate.

Byte Order

It is important to realise that nearly all serial communications systems transmit the least significant bit of each byte first at the physical layer. Ethernet supports broadcast, unicast, and multicast addresses. The appearance of a multicast address on the cable (in this case an IP multicast address, with group set to the bit pattern 0xxx xxxx xxxx xxxx xxxx xxxx) is therefore as shown below (bits transmitted from left to right):

```
0                          23 IP Multicast Address Group     47
|                          |<-------------------------------->|
1000 0000 0000 0000 0111 1010 xxxx xxx0 xxxx xxxx xxxx xxxx
|                                    |
Multicast Bit                        0 = Internet Multicast
                                     1 = Assigned for other uses
```

However, when the same frame is stored in the memory of a computer, the bits are ordered such that the least significant bit of each byte is stored in the right most position (the bits are transmitted right-to-left within bytes, bytes transmitted left-to-right):

```
0                          23                                47
|                          |                                 |
0000 0001 0000 0000 0101 1110 0xxx xxxx xxxx xxxx xxxx xxxx
     |                          <---------------------------->
Multicast Bit                   IP Multicast Address Group
```

1.12. CSMA /CD

The Carrier Sense Multiple Access (CSMA) with Collision Detection (CD) protocol is used t control access to the shared Ethernet medium. A switched network (e.g. Fast Ethernet) may use a fu duplex mode giving access to the full link speed when used between directly connected NICs, Switc to NIC cables, or Switch to Switch cables.

Receiver Processing Algorithm

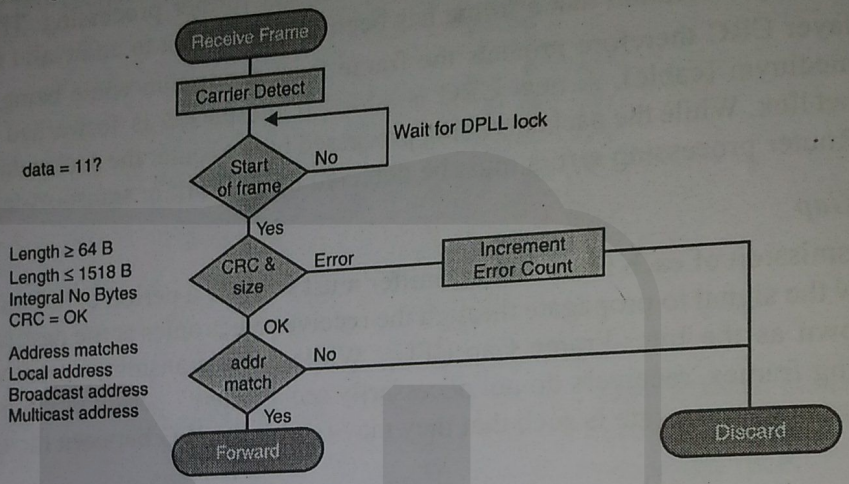

Fig. 1.13.

Runt Frame

Any frame which is received and which is less than 64 bytes is illegal, and is called a "runt". most cases, such frames arise from a collision, and while they indicate an illegal reception, they m be observed on correctly functioning networks. A receiver must discard all runt frames.

Giant Frame

Any frame which is received and which is greater than the maximum frame size, is called "giant". In theory, the jabber control circuit in the transceiver should prevent any node from generati such a frame, but certain failures in the physical layer may also give rise to over-sized Ether frames. Like runts, giants are discarded by an Ethernet receiver.

Jumbo Frame

Some modern Gigabit Ethernet NICs support frames that are larger than the traditional 15 bytes specified by the IEEE. This new mode requires support by both ends of the link to supp Jumbo Frames. Path MTU Discovery is required for a router to utilise this feature, since there is other way for a router to determine that all systems on the end-to-end path will support these lar sized frames.

A Misaligned Frame

Any frame which does not contain an integral number of received bytes (bytes) is also illegal. receiver has no way of knowing which bits are legal, and how to compute the CRC-32 of the fra Such frames are therefore also discarded by the Ethernet receiver.

Other Issues

The Ethernet standard dictates a minimum size of frame, which requires at least 46 bytes of data to be present in every MAC frame. If the network layer wishes to send less than 46 bytes of data the MAC protocol adds sufficient number of zero bytes (0×00, is also known as null padding characters) to satisfy this requirement. The maximum size of data which may be carried in a MAC frame using Ethernet is 1500 bytes (this is known as the MTU in IP).

A protocol known as the "Address Resolution Protocol" (arp) is used to identify the MAC source address of remote computers when IP is used over an Ethernet LAN.

Exception to the Rule

An extension to Ethernet, known as IEEE 802.1p allows for frames to carry a tag. The tag value adds an extra level of PCI to the Ethernet frame header. This increases the size of the total MAC frame when the tag is used. A side effect of this is that NICs and network devices designed to support this extension require a modification to the jabber detection circuit.

■■■

CHAPTER 2

The Network Laye

The network layer is a host to host delivery layer *i.e.* for carrying the packet from the source to destinatio It deals with deals with end-to-end transmission and is the lowest layer to do so. It is the third layer in t network hierarchy and provides service to the Transport Layer. It uses the services of Data Link layer to this. The unit of data that is handled by the network layer is known as a packet. To provide services to t transport layer, a unique address is required. This unique address is defined by the network layer and known as the logical address. The equivalent of network layer in TCP/IP reference model is known as t **Internet Layer.**

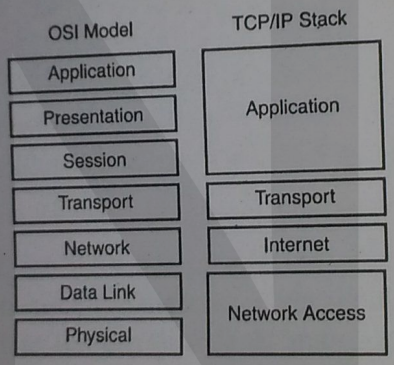

Fig. 2.1. Position of Network Layer in Layered Architecture

The main functions of this network layer are as follows:

(a) **Addressing:** To identify each device each device on the Internet uniquely, addressing required. The address used in the network layer should be unique and is known as logical address.

(b) **Internetworking:** There are many different types of networks and the network lay provides a logical connection between these different networks.

(c) **Routing:** In a network, there are multiple routes available from source to destinati andone of them is to be chosen. It is this network layer that decides the route to be take This is called as routing.

(d) **Packetizing Data:** Data at the network layer is handled in form of packets. The netwo layer encapsulates the packets received from upper layer protocol and makes new packe

This is called as packetizing. The size and format of packet varies from network to network, but it remains as one logical unit.

(e) **Fragmenting:** The datagram can travel through different networks. Each router decapsulates the IP datagram from the received frame. Then the datagram is processed and encapsulated in another frame.

(f) **Cost and Billing Information:** The network layer is responsible for maintaining a record of statistics related to data usage, data transmitted, etc. this information is required to be used for billing purposes.

To provide these services, the network layer is designed in such a way, that it works independently of underlying technology. Users of the service need not be aware of the physical implementation of the network.

2.1 NETWORK LAYER DESIGN ISSUES

The following design issues are related to the services offered by the network layer:

1. Providing a link between data link layer and the transport layer.
2. Providing Routing and Delivery services to end destination.
3. Maintaining accounting and Statistical Information for billing purposes.
4. Handling different packet formats
5. Handling different addressing schemes used in different networks
6. Providing error recovery at end user, if error is detected during transmission.
7. Providing Connection-oriented or Connectionless Services: different networks support different types of connections: connection-oriented or connection-less. The network layer should be capable of providing both types of services.
8. Providing Unique addressing to different nodes and sub-networks.

2.2 ADDRESS RESOLUTION PROTOCOL

ARP stands for Address Resolution Protocol. To reach to a destination, a packet has to travel through multiple networks and devices such as routers. At the network level, the hosts and routers are recognized by their IP addresses. The IP address is a universally unique address and no two hosts can have same IP address. Transmission from the Network layer is handed over to DataLink layer for transmission. But the DataLink layer does not recognize IP addresses. At the physical level, the IP address is not useful, so the hosts and routers are recognized by their MAC addresses. A MAC address is a local address which is unique locally but not globally. Therefore, both IP and MAC addresses are required for communication. It is possible that a physical network can have 2 different protocols at the network layer. Thus, for correct packet delivery, we need a 2-level addressing, namely IP addressing and MAC addressing. We should be able to map the IP addresses into a corresponding MAC address and vice-versa. This mapping can be static or dynamic:

Static mapping: The static mapping uses a table to store physical addresses corresponding to every IP address. This table is stored on every machine. If the machine knows the IP address of another machine then it can search for the corresponding MAC address in its table. The problem here is that the MAC

addresses can change. Thus, to implement static mapping, the static mapping table needs to be up periodically.

Dynamic mapping: In this type of mapping, we use a protocol for finding one address fro other. There are 2 protocols in use: ARP and Reverse ARP (RARP). ARP maps an IP address to a address whereas the RARP maps a MAC address to a given IP address.

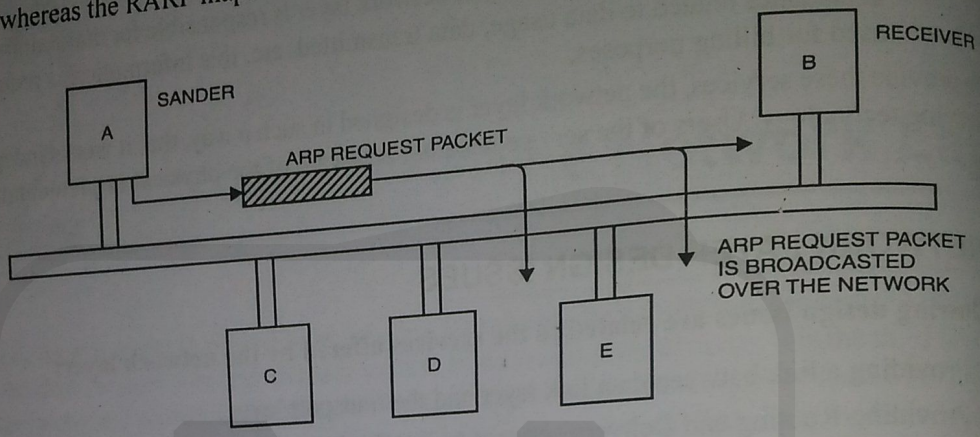

Fig.2.2. ARP request Broadcast

Working of ARP: ARP is used for associating an IP address to its MAC address. Every de also has a physical or MAC address, which is useful for identification at the Data Link layer. address is available in the Network Interface Card (NIC) of every device. Every device on a Net needs to have a NIC. The process of finding the MAC address of another host or network can be sum up as:

Step-1: The router or host-A, who wants to find the MAC address of some other router, sends an request packet. This packet contains the IP and MAC addresses of the sender and also t address of the receiver-B.

Step-2: This request packet is broadcasted over the network as shown in Fig. 2.2.

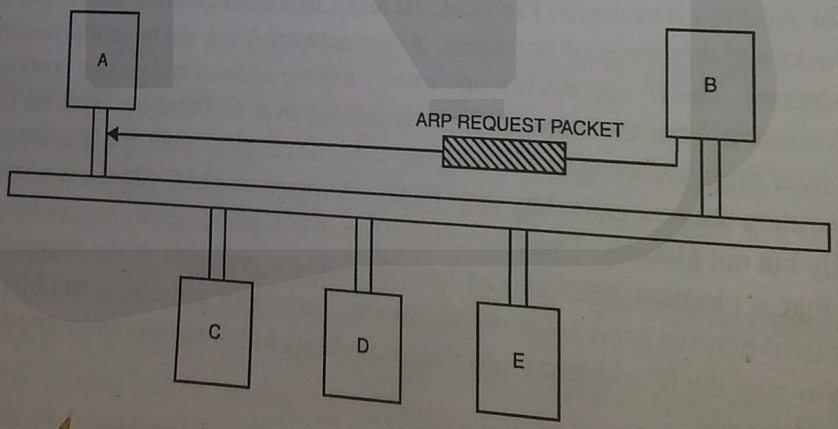

Fig. 2.3. ARP Response Unicast

addresses can change. Thus, to implement static mapping, the static mapping table needs to be updated periodically.

Dynamic mapping: In this type of mapping, we use a protocol for finding one address from the other.There are 2 protocols in use: ARP and Reverse ARP (RARP). ARP maps an IP address to a MAC address whereas the RARP maps a MAC address to a given IP address.

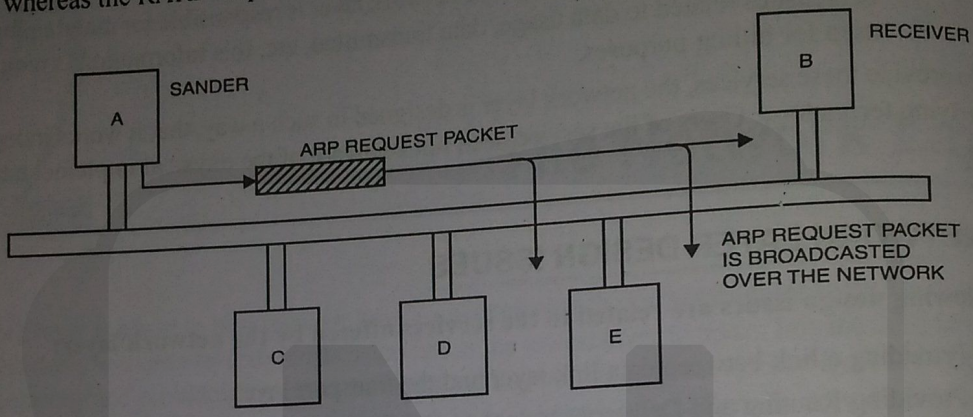

Fig.2.2. ARP request Broadcast

Working of ARP: ARP is used for associating an IP address to its MAC address. Every device also has a physical or MAC address, which is useful for identification at the Data Link layer. This address is available in the Network Interface Card (NIC) of every device. Every device on a Network needs to have a NIC. The process of finding the MAC address of another host or network can be summed up as:

Step-1: The router or host-A, who wants to find the MAC address of some other router, sends an ARP request packet. This packet contains the IP and MAC addresses of the sender and also the IP address of the receiver-B.

Step-2: This request packet is broadcasted over the network as shown in Fig. 2.2.

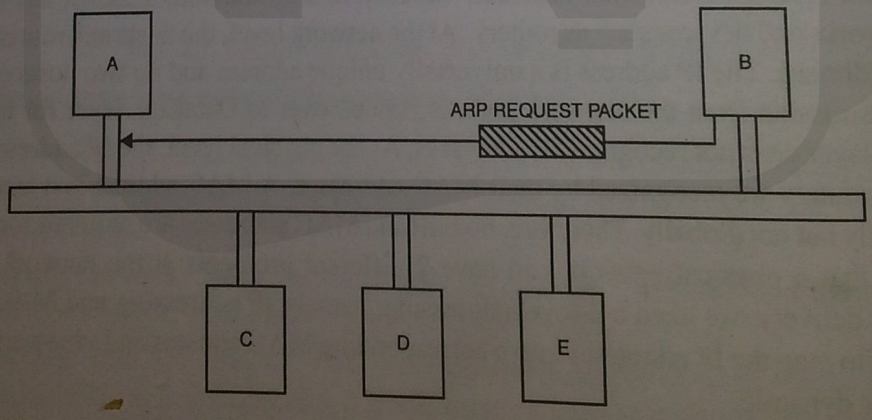

Fig. 2.3. ARP Response Unicast

Step-3: Every host and router on the network receives and processes the ARP request packet. All hosts except B discard the packet and only the intended receiver (B) recognizes its IP address in the request packet. The host B responds back by sending an ARP response packet.

Step-4: ARP response packet contains the IP and physical addresses of the receiver (B). This packet is Unicasted only to A using A's physical address. It is shown in Fig. 2.3.

2.2.1. ARP Packet Format

HARDWARE TYPE (16 BITS)		PROTOCOL TYPE (16 BITS)
HARDWARE LENGTH	PROTOCOL LENGTH	OPERATION REQUEST 1, REPLY 2
SENDER HARDWARE ADDRESS		
SENDER PROTOCOL ADDRESS		
TARGET HARDWARE ADDRESS		
TARGET PROTOCOL ADDRESS		

Fig. 2.4. ARP packet Format

1. **HTYPE (Hardware Type):** It specifies a hardware interface type. It contains a value of 1 for Ethernet. It is 16 bits field that defines the type of network on which ARP is being run. ARP can run on any physical network.
2. **PTYPE (Protocol Type:** It specifies the type of high level protocol address. The sender has supplied for IP address 0800 H (value). It is a 16 bit field that is used to define the protocol using ARP. **ARP can be used with any higher-level protocol such as IPv4.**
3. **HLEN (Hardware length):** It is an 8-bit field used to define the length of the physical address in bytes. Like for Ethernet, this value is 6.
4. **PLEN (Protocol length):** It is 8-bit long field. It defines the length of the IP address in bytes. For IPv4, this value is 4.
5. **OPER (Operation):** It is a 16 bit field that defines the type of packet. It specifies the ARP request. They are—ARP response and ARP reply.
6. **SHA (Sender Hardware Address):** This field defines the physical address of the sender. The length of this field is variable.
7. **TPA (Target Protocol Address):** This field defines the logical address of the target. It is a variable length field.
8. **THA (Target Hardware Address):** It defines the physical address of the target. It is a variable length field. For the ARP request packet, this field contains all zeros because the sender does not know the receiver's physical address.
9. **TPA (Target Protocol Address):** This field defines the logical address of the target. It is a variable length field.

2.3 REVERSE ADDRESS RESOLUTION PROTOCOL

Sometime a situation arises that a node knows its physical address but does not know its logical or IP address. This situation arises in some special cases:

- like Booting of a diskless station
- situations where IP addresses are assigned dynamically and not statically.

In these situations a machine knows can know its physical address by looking at its Network Interface Card, but does not know its logical address. A host needs to know its logical address to communicate with other hosts through a IP datagram. RARP has been designed to address this problem of obtaining IP address from the physical address for a device. The requesting machine acts like a RARP Client and creates a RARP request. This request is broadcasted on the LAN. A machine on the LAN which knows all IP addresses acts like an RARP Server. It responds by sending the IP address matching with the physical address. This IP address is returned in the RARP reply which is unicasted to the RARP Client.

The drawback of this method is that the RARP Request is broadcasted on the network and it cannot pass on to other networks. Therefore, an RARP Server is required to be implemented on each different network. An alternate to RARP is the Bootstrap (BOOTP) or the DHCP Protocol, which run at the application layers.

2.4 INTERNET CONTROL MESSAGE PROTOCOL (ICMP)

Internet communication mostly takes place in form of IPV4 datagrams. Internet protocol is not designed to be reliable whenever there is some problem/error in data transmission, ICMP can be used for error reporting. ICMP is an error reporting protocol, where messages are sent to source IP address to report about some error in delivery. ICMP messages can be sent & received by any IP N/W device. ICMP is used by pouters, hosts or other devices to communicate errors or updates to other devices. They are also used for diagnosis & troubleshooting purposes. ICMP message are transmitted in form of datagrams, with a 8 byte long header. ICMP messages cuaibe the entire IP heater of the original message, which was failed. ICMP packets are IP Packets with ICMP in the IP data portion. ICMP is formally defined in RFC 792-ICMP messages are aimed at providing the feedback but these is no guarantee that a datagram will be delivered.

ICMP header is 8 bytes long. The first 4 for bytes always have the same meaning & the next 4 bytes vary according to ICMP message type. The format of ICMP header is:

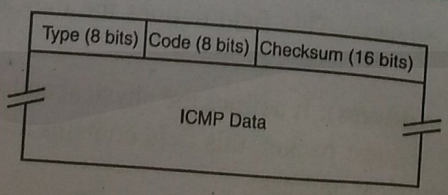

Fig. 2.5. The ICMP header

1. The type field identifies the type of message. These are many types of ICMP messages.

2. Code: Some ICMP message types have different codes associated with them. These codes are used to identify the sub-type of a message. Not all message types have sub-types associated with them.
3. Checksum: The checksum field of ICMP message contains error checking data calculated from the header & data. This field is 16 bits long.
4. ICMP Data: This is a variable length field. I contains data specific to a message type & code typefields.

ICMP TYPE NUMBERS

The Internet Control Message Protocol (ICMP) has many messages that are identified by a "type" field.

Type	Name
0	Echo Reply
1	Unassigned
2	Unassigned
3	Destination Unreachable
4	Source Quench
5	Redirect
6	Alternate Host Address
7	Unassigned
8	Echo
9	Router Advertisement
10	Router Selection
11	Time Exceeded
12	Parameter Problem
13	Timestamp
14	Timestamp Reply
15	Information Request
16	Information Reply
17	Address Mask Request
18	Address Mask Reply
19	Reserved (for Security)
20-29	Reserved (for Robustness Experiment)
30	Traccroute
31	Datagram Conversion Error
32	Mobile Host Redirect
33	IPv6 Where-Are You

34	IPv6 I-Am-Here
35	Mobile Registration Request
36	Mobile Registration Reply
37	Domain Name Request
38	Domain Name Reply
39	SKIP Algorithm Discovery Protocol
40	Photurius
41-255	Reserved

Some Common ICMP Message Types are:

1. Echo Request and Echo Response: ICMP is often used to test the connectivity between devices. The ping command uses the ICMP protocol. The ping is a command-line utility to check the connectivity between two devices. The ping command sends an IP datagram to an IP address and the destination responds by sending a response datagram. Ping command uses Internet Control Message Protocol (ICMP) Echo Request and Echo Reply.
2. Destination Unreachable: ICMP returns a Destination Unreachable message to the source IPv4 address when a datagram that cannot be delivered, is received by a router.
3. Time Exceeded: This message is sent to the source IP if a datagram is discarded because Time-to-Live (TTL) value reaches zero. This happens when the number of hops exceeds the maximum value.
4. Source Quench: If a device is sending large amounts of data to another remote device, the volume can flood the router with data. The router can use Internet Control Message Protocol (ICMP) to send a Source Quench message to the source IPv4 address to ask it to slow down the rate at which it is sending data.

2.5 ROUTING BASICS

Routing is defined as the process of moving information from a source to a destination. This information is moved in logical Units called packets. The end devices may be connected within a same network or they may be in different networks. So addressing different packet formats and different addressing schemes is all a part of routing. And the devices that are used to connect two or more networks are known as routers. They provide the inter-networking between different types of networks. They consist of a combination of hardware and software.

Routers use logical and physical addressing to connect two or more logically separate networks. They accomplish this connection by organizing the large network into logical network segments called as subnets. Each of the subnet is given a logical address. This allows the networks to be separate but still exchange data. Data is grouped into packets or blocks of data. Each packet has a physical device address as well as logical network address. The role of routers is not only to establish paths between two end pints, but to do it efficiently. For this, the routers need to calculate optimal path to a workstation or computer.

2.5.1 How a Routers works

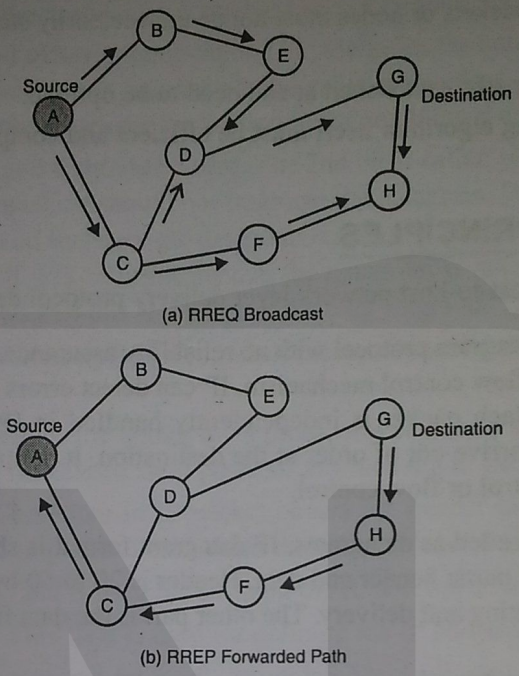

Fig. 2.6. The Routing Process

Routers maintain the information about the network in form of tables called Routing Tables. A routing table maintains a row for each destination, the path or next node to be taken and a metric associated with the path. Using the next node information, the router knows which path a packet must be forwarded to. When a packet arrives at a router for some destination, router simply searches its routing table to determine if it has a path to that destination. If it finds one, it simply forwards the packet to Next Node. If no match is found, it means that no path to destination exists, and a new path is to be searched. In this situation, a RouteRequest message is generated by the node and it broadcasts the message to all its neighbours with a timestamp. Intermediate node simply forward the message to other neighbouring nodes if they do not have any valid path. When this RouteRequest reaches the destination or an Intermediate node which knows the path to destination, it replies by creating a RouteReply message. This way the RouteReply traverses back to Source and path is established. The source and all Intermediate nodes make an entry in their respective Routing Tables. When multiple paths are discovered, the metrics associated with each route are also recorded into the tables. The metric is a value which describes the path such as delay, hops, bandwidth, jitter etc.

2.5.2 Routing Algorithms

A routing algorithm is a part of network layer software. It is responsible for deciding the output path over which the packet must be sent. Many different and diverse Routing Algorithms have been developed for dealing with different scenarios. Each routing algorithm tries to find an optimal path among the many alternates available.

The following are the properties of any routing algorithms:

1. It's correctness: The algorithm must produce an efficient path at the end of routing process.
2. Its robustness: The routing algorithm must be able to work even correctly even under different kinds of failure.

3. It's stability: The paths established must be stable and loop free.
4. It's fairness: Some packets or nodes must not be unnecessarily starved. The algorithm must be as fair as possible.
5. It's optimality: The paths established at end need to be optimal.
6. Efficient: The routing algorithm itself must be efficient and compute the paths with minimum overheads.

2.6 IPV4 ROUTING PRINCIPLES

Internet Protocol or IP is a host-to-host network layer delivery protocol designed for the internet.

It is a connectionless datagram protocol with no reliability assurance. It is unreliable as it does not provide any error control or flow control mechanism. IP can detect errors in transmission, but discards the message if corrupted. Each packet is independently handled in IP. These packets can follow independent paths and even arrive out of order at the destination. It is upto the higher layer protocols like TCP, to handle error control or flow control.

Packets in IP layer are called as datagrams. IP datagram format is shown below. A datagram is a variable length packet with 2 parts: header and data. Header is 20 to 60 bytes in length. It contains the information necessary for routing and delivery. The other part is the data field that is of variable length.

2.6.1 IPv4 Datagram Header

It contains the routing information and control information associated with datagram delivery as shown below:

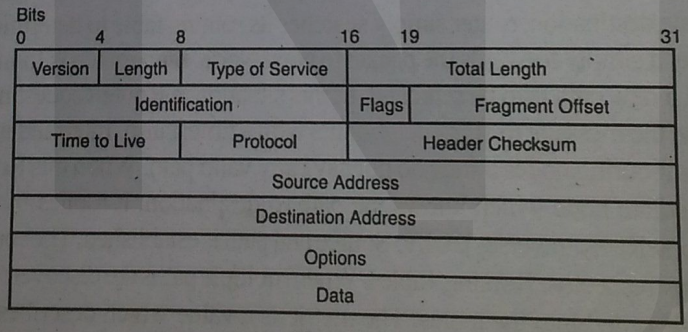

Fig. 2.7.

1. **VER (version):** This field defines the version of IP. Current version of IP is IPv4 and the latest version is IPv6. it is a 4-bit long field.
2. **HLEN (Header Length):** This field defines the length of the datagram header in 4-byte word. Its value must be multiplied by 4 to give the length in bytes.
3. **Differential Services (DS):** This field defines the class of the datagram for quality of service (QoS). During heavy traffic, a three way trade-off exists between low delay, high reliability and throughput.
4. **Total Length:** It gives the total length of the IP datagram that includes the length of the header as well as the data field. This field is of 16-bits. Hence, the total length is limited to $(2^{16} - 1) =$ 65,535 bytes. Out of this, 20 to 60 bytes are the header and the remaining are the data. All hosts

must be ready to accept datagram of upto 576 bytes. The hosts should send the datagrams larger than 576 bytes only if the destination is ready to accept larger datagrams.

5. **Identification, flag and offset:** Identification field identifies the datagram originating from the source host. When a datagram is fragmented, the value in the identification field is copied into all fragments. This number helps the destination in reassembling the fragments of the datagram. The 1st bit is reserved and it should be zero. The 2nd bit is called as 'Do Not Fragment' bit. If this bit is '1' then the machine should not fragment the datagram. But if the value of this bit is 0 then the machine should fragment the datagram if and only if required. the 3rd bit is called as 'More Fragment Bit'. If it is '1' it means that the datagram is not the last fragment but if its value is '0' then it shows that this is the last or the only fragment.

   Time to Live: It is an 8-bit long field. It controls the maximum number of routers visited by the datagram.

7. **Protocol:** This field defines the higher level protocol which uses the services of the IP layer. This field specifies the final destination protocol to which the IP datagram should be delivered.
8. **Header Checksum:** Checksum in IP packet covers the header only. As some header fields change, this field is recomputed at each point the Internet header is processed.
9. **Source Address:** This field is used for defining the IP address of the source.
10. **Destination Address:** This field is used for defining the IP address of the source.
11. **Options:** They are not required for every datagram. They are used for network testing.

2.6.2 IP Addressing

IP addressing can be either classful or classless:

1. **Classful addressing**

   Classic TCP/IP addressing architecture divides all possible IP addresses into classes based on the most significant bits of the address. The class determines what part of the address designates the network and what part designates the host. The three main address classes are Class A, Class B and Class C.

   Class A addresses have a binary zero (0) as the most significant bit of the address. The address range is from 0.0.0.0 to 127.255.255.255. **The network portion is the most significant byte** of the address and the host portion is the three byte remainder. Class A network addresses are intended for networks that have more than 65,534 hosts on a single network.

   Class B addresses have binary One Zero (10) as the most significant two bits of the address. The address range is from 128.0.0.0 to 191.255.255.255. **The network portion is the most** significant two bytes of the address and the host portion is the two byte remainder. Class B network addresses are intended for networks that have from 255 to 65,534 hosts on a single network.

   Class C addresses have binary One One Zero (110) as the most significant three bits of the address. The address range is from 192.0.0.0 to 223.255.255.255. **The network portion is the most significant three bytes of the address and the host is one byte remainder.** Class C network addresses are intended for networks that have fewer than 255 hosts on a single network

Class D: first four bits of the first octet in Class D IP addresses are set to 1110. Class D has IP address rage from 224.0.0.0 to 239.255.255.255. Class D is reserved for Multicasting. In multicasting data is not destined for a particular host, that is why there is no need to extract host address from the IP address, and Class D does not have any subnet mask.

Class E addresses: This IP Class is reserved for experimental purposes only for R&D or Study. IP addresses in this class ranges from 240.0.0.0 to 255.255.255.254.

2. **Classless Addressing**

Although the number of actual devices connected to Internet is much lower than the address space of 4 billion, the address depletion has taken place due to flaws in the classful addressing scheme. Class A and Class B addresses are the most affected. An alternate to overcome this limitation is to use classless addressing. **In this technique, there are no classes but the addresses are still generated in blocks. In this scheme, when an entity needs to be connected to the internet, a block range of addresses** is granted to it. The size of this block granted is equal to actual requirement, it is not fixed like classful addressing. It means that for a small entity like household only one or two addresses can be provided, but for a lager organization, even thousands of addresses can be allotted.

The following rules are followed while allotting classless address blocks:

(*a*) The address in a block must be continuous.

(*b*) The number of addresses in a block should be a power of 2.

(*c*) The first address should be evenly divisible by the number of addresses.

Classless addressing treats the IP address as a 32 bit stream of ones and zeroes, where the boundary between network and host portions can fall anywhere between bit 0 and bit 31. The subnet mask associated with an address determines the network address. A subnet mask is used locally on each host connected to a network, and masks are never carried in IPv4 datagrams. All hosts on the same network are configured with the same mask, and share the same pattern of network bits. The number of 1's in the subnet mask determines the network address.

2.7 CLASSIFICATION OF ROUTING ALGORITHMS

Routing algorithms can be classified either according to their adaptation ability (static or dynamic) or according to their range of operation (Intra and Inter Domain).

2.7.1 Acccording to their adaptation ability, the routing algorithms are put under two categories:

(*a*) **Non-adaptive algorithms:** In these types of algorithms, the routing decision is not based on the measurement or estimation of current traffic and topology. However, the choice of the route is done in advance, off-line and it is downloaded to the routers. **This is also called as static routing. Examples of such algorithms include OSPF, BGP**

(*b*) **Adaptive Routing:** In these types of algorithms, the routing decision can be changed if there are any changes in the topology or traffic. The paths are variable and can change during course of time. **This is also called as dynamic routing. Examples of such algorithms include Distance vector Routing and Link State Routing.**

2.7.2 On the basis of their range (domain) of operation, routing algorithms are classified as Intra and Inter Domain Routing.

The Internet is a large network consisting of many networks. Several heterogeneous networks make up the Internet. For this we divide the Internet into different Autonomous Systems (AS). An Autonomous System is a collection of networks, routers and other devices that fall under the authority of a single administration. The rules that are used for communication are known as protocols. A single Autonomous Systems generally uses a single or a combination of some protocols. The routing protocols fall into two categories:

(*a*) **Intra-Domain Routing Protocols.** The protocols used within a single Autonomous System are known as Intra-Domain or Interior Routing Protocols (IRP). Link State Routing and Distance Vector Routing are the two main classes of methods used in an Intra-domain system.

(*b*) **Inter-Domain Routing Protocols:** The protocols used to connect two or more different Autonomous Systems are known as Inter-Domain Routing Protocols or Exterior Routing Protocols (ERP). Path Vector Routing Protocol is used in an Inter-Domain system.

Now we discuss two important routing methods used in an Intra-domain environment: The Link State Routing and Distance Vector Routing methods.

2.8 LINK STATE ROUTING

Principle: Link state routing must perform 5 basic router operations, as follows:

1. Each router should discover its neighbours and obtain their network addresses.
2. Then it should measure the delay or cost to each of these neighbors.
3. It should construct a packet containing the network addresses and the delays of all of the neighbors. These packets are called link state advertisements (LSAs).
4. Send LSA packet to all other routers in the network.
5. Each router maintains a database of all received LSAs (topological database or link state database), which describes the network has a graph with weighted edges
6. Each router uses its link state database to run a shortest path algorithm (Dijikstra's algorithm) to produce the shortest path to each network

The complete topology and all delays are experimentally measured and this information is conveyed to each and every router. A shortest path algorithm like Dijkshtra's algorithm can be used to find the shortest path to every other router in case of Link State Routing. Link state routing is very commonly used. OSPF and the IS-IS protocol or Intermediate System—Intermediate System use link state algorithm. It is mainly used in Internet backbones and cellular systems.

LSR Features

- In link state routing, each node has a complete map of the topology
- LSAs are flooded to all nodes in the network.
- Updates are sent only when some changes occur in a neighbour (change in cost or the node goes down)
- If a node fails, each node can calculate the new route
- Difficulty: All nodes need to have a consistent view of the network

2.9 DISTANCE VECTOR ROUTING

Distance vector protocols (a vector contains both distance and direction) determine the path to remote networks using hop count as the metric. The Distance vector Routing method has the following steps:

- Initially every node discovers its neighbour nodes and the cost associated with them
- This information is stored in their routing tables.
- A node exchanges this routing table with its immediate neighbours only.
- If a new destination is received by a node in this exchange, it is also added to its routing table
- The cost of new path is calculated as the sum of cost from the current node to its neighbour+ the cost of neighbour to the destination.
- Periodic updates are sent at regular intervals.

DVR Features

- With distance vector routing, each node has information only about the next hop.
- As information propagates during subsequent cycles, it becomes known to every router.
- Periodic updates are sent at a set interval.
- Updates are sent to neighbour nodes only.
- Bellman Ford algorithm is generally used to find shortest routes from the received routing tables.
- Distance vector routing makes poor routing decisions if directions are not completely correct.

Examples of DVR include RIP and IGRP.

2.10 INTRA-DOMAIN ROUTING PROTOCOLS

Unlike Inter-Domain protocols, these protocols operate within a single AS. Therefore, their reach is limited to shorter geographical areas, when compared with **Inter-Domain Routing Protocols. We now discuss some important Intra-domain routing protocols:**

2.10.1 Open Shortest Path First protocol (OSPF)

OSPF is an Intra-domain Protocol used to build shortest paths in a network. It is a Link State Routing Protocol. It uses the Dijksttra's Algorithm to build shortest-paths in a H/W. The steps of operation are:

1. Each node discovers its neighbours and the cost to reach neighbours is determined. This information is stored in form of packets called Lind State Advertisement (LSA).
2. Forwarding of LSAs to all neighbour nodes by flooding.
3. Each node collects all received LSAs to construct N/W topology.
4. Dijkstra's Shortest Path algorithm is applied to the n/w topology to arrive at Shortest Paths to each node.

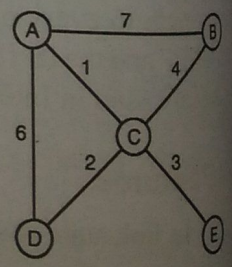

Fig. 2.8. A Sample Network

Example: Consider the following n/w which is built using the LSAs received.

Every node will apply Dijkstra's algo on same graph to calculate Shortest Paths. The Shortest Paths for every node will be different due to their respective positions.

Shortest Paths for Node A:

Node A determines all its neighbours & puts them in a queue arranged by distance from Node A. The nodes which are unreachable their distance is marked as ∞.

Step 1

A	C	D	B	E
0	1	6	7	∞

Node A is Made Permanent & taken out from the queue.

Step 2

Permanent : A

C	D	B	E
1	6	7	∞

Node C is Made Permanent & taken out from the queue. Revised distances are calculated from node C.

Steo 3

Parmanent : A, C

D	E	B
3	4	5

Node D is Made Permanent & taken out from the queue.

Step 4

E	B
4	5

Node E is Made Permanent & taken out from the queue.

Step 5

B
5

Lastly B is Made Permanent & taken out from the queue. Now the queue becomes empty. Thus the Shortest Paths as determine are stored in a routing table.

Destination	Cost	Next
A	0	–
B	5	C
C	1	C
D	3	C
E	4	C

Shortest Path Tree built by Node A

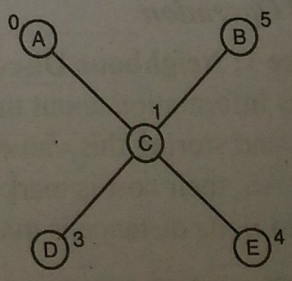

Fig. 2.9. Shortest Path tree built by Node A

2.10.2 OSPF for IPv4 and OSPF for IPv6

OSPF has gone through changes over the years and the protocol has been adapted to work with IPv6. The earlier version of OSPF was OSPFv2. As organizations are transforming to IPv6, there is need to upgrade OSPFv3, which is compatible with both IPv4 and IPv6. The IETF modified OSPF through a series of RFCs and was updated through RFC 5340. Here are some of the differences between OSPFv2 and OSPFv3.

OSPFv2	OSPFv3
1. OSPFv2 is for IPv4-only.	OSPFv3 can be used for IPv4 or IPv6
2. LSA packet format is different	New types of LSA packets have been introduced
3. OSPFv3 runs per-subnet	OSPFv3 runs per-link basis rather than per-subnet
4. OSPFv3 uses different flooding scope bits	There are separate scopes for flooding LSAs: **Link-local scope** **Area scope** **AS scope**
5. We can use only one instance on one link to create a single link in more than one	We can use multiple instances on the same link area
6. OSPFv2 authentication is achieved by implementing a shared secret and MD5 HMAC supported as part of the OSPFv2 protocol	OSPFv3 does away its own support for authentication and uses the IPsec framework offered by IPv6
7. IP addressing is not separate of calculating the Shortest Path tree	Separation of IP addressing from the calculation of the Shortest Path tree. Because of this separation, adding or modifying IP subnets within the OSPF domain will not affect the integrity of the Shortest Path trees.

2.10.3 Routing Information protocol (RIP)

It is a Distance vector Routing Protocol. The basis for RIP is the Bellman-Ford Algorithm. The algorithm works in a distributed way to build shortest paths between different nodes. Every node maintains a routing table containing address of different nodes, their cost and the next Hop. The next hop field tells the next node where data must be transmitted to reach a particular node.

RIP Operation

Phase 1: Neighbour Discovery: Initially, every node only knows about its immediate neighbours and stores information about them in their routing tables. Their cost is calculated by sending a message to them and storing this value in its routing table. Nodes which are not directly reachable, but within the same AS, their cost is marked as infinity (∞). Consider a network as below. Since A does not know any path to E, its distance is maintained as ∞.

Phase 2 (Advertising): In this phase, immediate neighbours exchange their routing tables to know about other nodes in the network. Paths are re-computed, if an alternate path becomes available during an advertisement exchange. The cost of new path is calculated as the sum of cost from the current node to its

neighbour+ the cost of neighbour to the destination. This old path is updated with this new path in two situations:

1. If the new path is received from a new Next Node and its cost is less then the previous cost. In case when the cost is equal or greater, the old path is retained.
2. If the path is received from the same Next Node and there is a change in cost (increase or decrease), the path is always updated. Even if the cost increases, this path is updated because this new message reflects the current network status. It is possible that some links were lost and therefore the cost may increase also.

Example: Consider the following topology

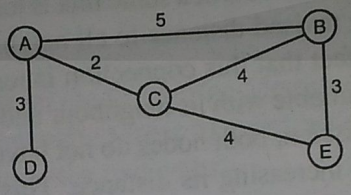

Fig. 2.10

To	Cost	Next
A	0	A
B	5	B
C	2	C
D	3	D
E	∞	—

Initial Routing Table of Node A

To	Cost	Next
A	2	A
B	4	B
C	0	C
D	∞	—
E	4	E

Initial Routing Table of Node C

Updation of A's table on receiving C's table

To	Cost
A	2
B	4
C	0
D	∞
E	4

Table Received from C

+

To	Cost	Next
A	0	A
B	5	B
C	2	C
D	3	D
E	∞	—

A's Original table

To	Cost	Next
A	0	A
B	5	B
C	2	C
D	3	D
E	6	C

A's Updated Table

The Count to Infinity Problem

RIP suffers from this problem. Problem occurs when some link is lost and its neighbouring node discover it and before it can communicate the revised distance to all neighbours, the neighbour node advertises the previous path, unmindful of the fact that this connection is lost. In this case, the neighbour of lo link thinks that an alternate path is available with its neighbour and reflects it into its routing table. Now when a data packet arrives for transmission, both nodes do not have a path to the lost link and they keep passing the packet to each other and increasing its distance. This problem is resolved only when the distance increases to infinity.

Solution

The solution to Count to Infinity problem is Split Horizon and Poisson reverse. Here when a node advertises its routes to a neighbour which is also the Next Node, it does not pass the actual cost value but instead replaces the cost value with a special marker to denote that the cost is what I know from you. Therefore, the neighbour will come to know that this advertisement is not a new path, but it already had

2.10.4 Interior Gateway Routing Protocol

IGRP is a distance Vector routing Protocol designed to be used in a Gateway for Connecting to many difference Networks has been designed to handle larger & more complex Networks. Some features of IGRP are:

1. Dynamic Routing, fast response to Network changes.
2. Stable routing ever in complexNetworks Networks.
3. Low overheads.
4. Division of load among different parallel routes for better Network characteristics.
5. Traffic load & error rates are taken into consideration to decide the flows.

IGRP is used in gateways connecting several different Networks. If a packet arriving at a gateway is for a Network connected to the gateway, the packed is forwarded to the Network & if the packet is for a for-away Network, then the gateway forwards the packets to another gateway on way to the Network.

IGRP Operation

Step 1 (Initialization): When a gateway is turned on, its routing table is initialized. A description be provided like delay values, bandwidth, etc.

All gateway are programmed to advertise their routing tables with other gateways at regular intervals.

Step 2 Advertising: In the second phase each gateway processes the routing table that it receives from other gateways & uses it to modify its own routing table. In this way the gateway comes to know

about networks connected to other gateways. The metric information obtained from other gateway help in optimizing the path. Selection, A single Composite Metric is used to evaluate different links. The formula for calculating this Composite Metric M is:

$$M = \left(\frac{k_1}{Be} + \frac{K_2}{Dc}\right) . r$$

Where K1 & K2 = Constants
Be = effective bandwidth
Dc = composite delay
r = fractional reliability (Percentage of Successful Packet Transmissions)

Step 3 Optimization: Based on the Composite Metrics above, a link is evaluated in terms of its quality. If two links with equal Composite Metric to a destination Network are available then IGRP will split the traffic almost equally among the two paths.

Example: Consider a Network as below:

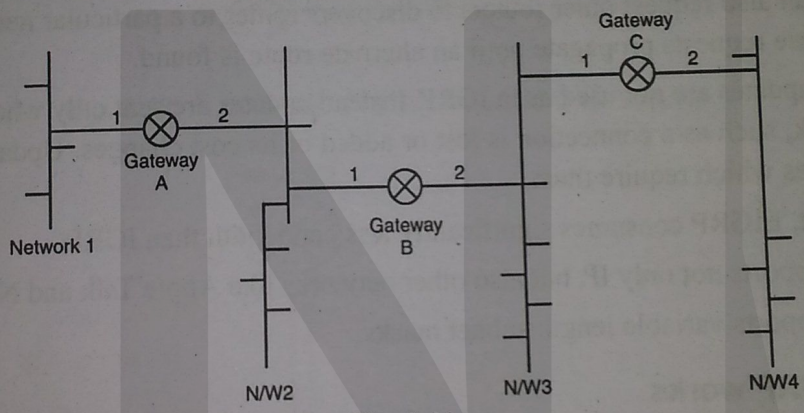

Fig. 2.11

Network	Gateway	Interface
1	None	1
2	None	2

Initial Routing table of Gateway A

Network	Gateway	Interface
2	None	1
3	None	2

Network	Gateway	Interface
4	None	2
4	None	2

Initial R.T. of Gateway B

When gateway B receives the R-Tables of A & C it updates & creates the following Routing table by Merging:

Network	Gateway	Interface
1	A	1
2	None	1
3	None	2
4	C	2

Final updated R.T. of Gateway B.

2.10.5 Enhanced IGRP (EIGRP)

Cisco introduced an enhanced version of IGRP that combines the advantages of link state protocols and distance vector protocols. Enhanced IGRP incorporates the Diffusing Update Algorithm (DUAL) developed at SRI International. Enhanced IGRP includes the following features:

1. Enhanced IGRP and DUAL to achieve fast convergence. In EIGRP, a router stores its neighbours tables also, so that it can quickly adapt to the network.
2. EIGRP can also request other routers to discovers routes to a particular network, if it does not exist. These requests propagate until an alternate route is found.
3. Periodic updates are not used as in IGRP. Instead updates are sent only when there is a change in network, such as a connection is lost or added or its cost changes. Updates are sent only to those nodes which require them.
4. As a result, EIGRP consumes significantly less bandwidth than IGRP.
5. EIGRP supports not only IP, but also other networks like Apple Talk and Novell Net Ware.
6. EIGRP supports variable length subnet masks.

2.10.6. How EIGRP works

When a router discovers a new neighbor, it records the neighbor's address and interface as an entry in the *neighbor table.* EIGRP uses Neighbour Discovery scheme. Every router periodically sends Hello packets to all nodes connected to it. As long as the packet is received, it is assumed that these nodes are alive. The DUAL modeule tracks all routes advertised by all neighbors. The metric advertised by all routers is used by DUAL to select efficient loop free paths. DUAL selects routes to be inserted into a routing table based on feasible successors. A successor is a neighboring router used for packet forwarding that has a least cost path to a destination. When a neighbor discovers a change in metric or when a topology change occurs, DUAL tests for feasible successors. If one is found, DUAL uses it to avoid recomputing the route unnecessarily, but when now successors are found, a recomputation is done among the advertising neighbours.

Example: Consider a network consisting of five routers as below in Fig. 2.12. For the destination N,

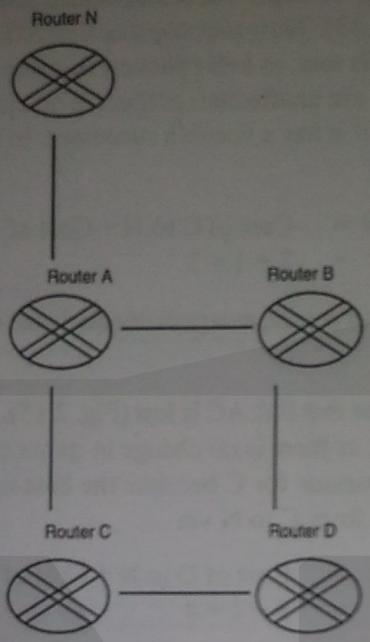

Fig. 2.12: A sample Network with 5 Routers

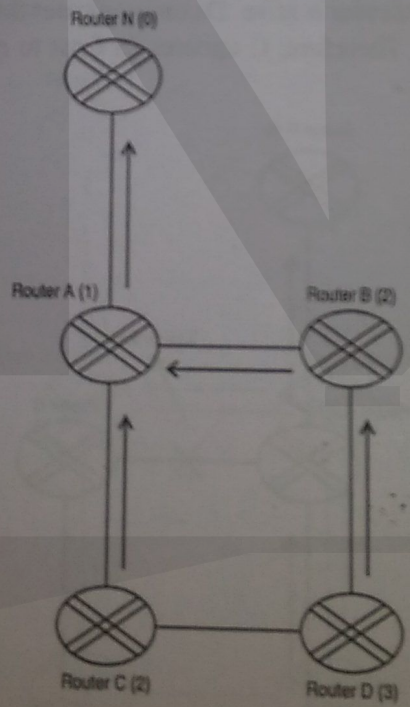

Fig. 2.13: Using EIGPR, successor nodes are marked.
Shortest paths are calculated (written in brackets)

Router A discovers N and advertises it to other routers. Router A marks router N as its successor. As the successive advertisements reach other routers, each router marks its successor node and records

the best distance in terms of Number of hops. Successor nodes have been marked by arrows and the distance is written in brackets (Fig. 2.13). Now suppose that link AB is lost after some time. Then router B, whose successor is A discovers this loss, as hello packets to A are lost. Now B advertises this change to all its neighbours. Router A and C are unaffected, as there is no change in their successors. But router D notices this change and discovers if it has a feasible successor. In this case, it discovers that C has an alternate path with a cost of 1.

Thus the total cost from D to N = Cost of C to N + Cost of D to C
= 2 + 1 = 3

Thus D relaces B with C as its successor, as a path of equal cost exist. Therefore, no re-computation is required. (Fig. 2.14)

Re-Computation: Now suppose that link AC is lost (Fig. 2.15). The successor of A i.e., C discovers this loss. B is unaffected by this loss, as there is no change in its successor. C advertises this information to its neighbours. Now D is not successor for C because the cost required to reach N via D is greater than the current value of 2. i.e. cost from C to N via

D = Cost of D to N + Cost of C to D
= 3 + 1 = 4

But this cost is greater than the current cost of 2 from C to N.

Now C must perform a route computation for destination N Fig. (2.16). It sends a query to router D. D need not re-compute, as its successor is alive. D communicates this cost to C. C comes to know that there is only one path to N via D. Therefore, C updates its Cost to reach N as 4 and also updates its successor to D.

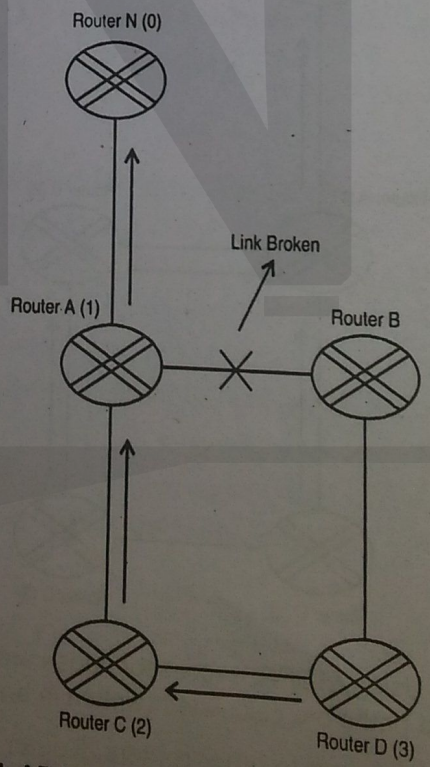

Fig. 2.14: Link AB is broken Node D finds a new Successor C with a total cost of 3. Here no re-computation is required.

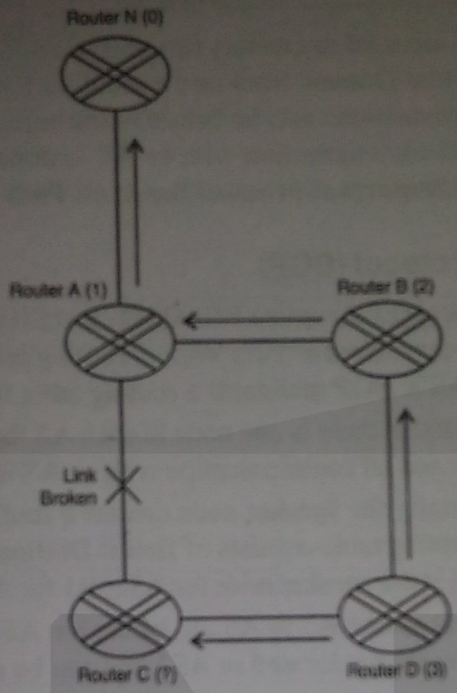

Fig. 2.15: Link AC is broken. Now a re-computation is required because C cannot find an alternate Successor.

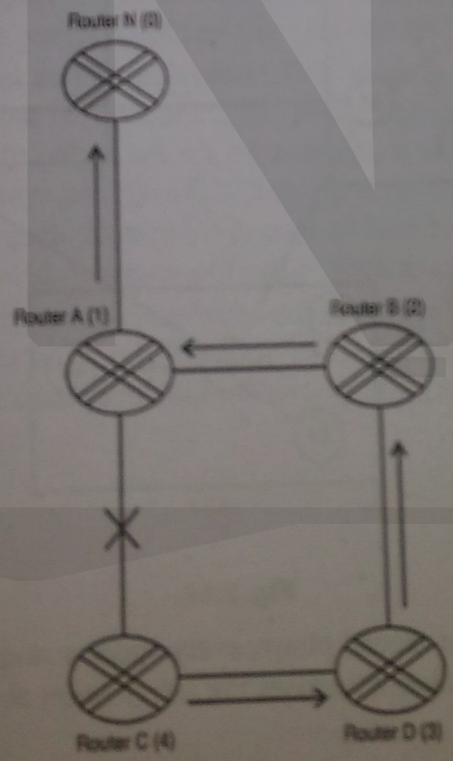

Fig. 2.16: A re-computation is performed by router C after link AC is lost. C updates its cost to reach N as 4 and its Successor as D.

2.11 INTER-DOMAIN ROUTING PROTOCOLS

Inter Domain Routing Protocols are used to connect two or more different Autonomous Systems. Path Vector Routing is an important Inter-Domain Routing Protocol. In Path Vector Routing, a special node is designated in each AS. This special node acts on behalf of the entire AS and is known as the speaker node. In Path Vector Routing, all communication with an AS is done through this speaker node only. **Border Gateway Protocol is an important protocol based on Path Vector Routing.**

2.11.1 Border Gateway Protocol (BGP)

BGP is an exterior gateway protocol (EGP) unlike RIP, OSPF, and EIGRP. The current BGP protocol is BGP Version 4 (BGPv4). BGP is considered a "Path Vector" routing protocol. BGP has been designed to route packets between different AS's. BGP maintains a routing table based on shortest AS Path.

BGP Operation: We assume that there is one node in each AS that acts on behalf of the entire AS. This node is called a speaker node and all communication with an AS is done through this speaker node.

Phase 1, Initialization: Initially the speaker node creates a routing table containing information of all nodes within its AS. This routing table consists of fields: Destination and Path, among other such as cost, hops, etc. For example, A1 is the speaker node for AS1, B1 for AS2, C1 for AS3 and D1 for AS4.

A1 creates an initial table that shows A1 to A5 are located in AS1 and can be reached through it. B1 creates a table that shows B1 to B4 are located in AS2 and can be reached through B1 and so on.

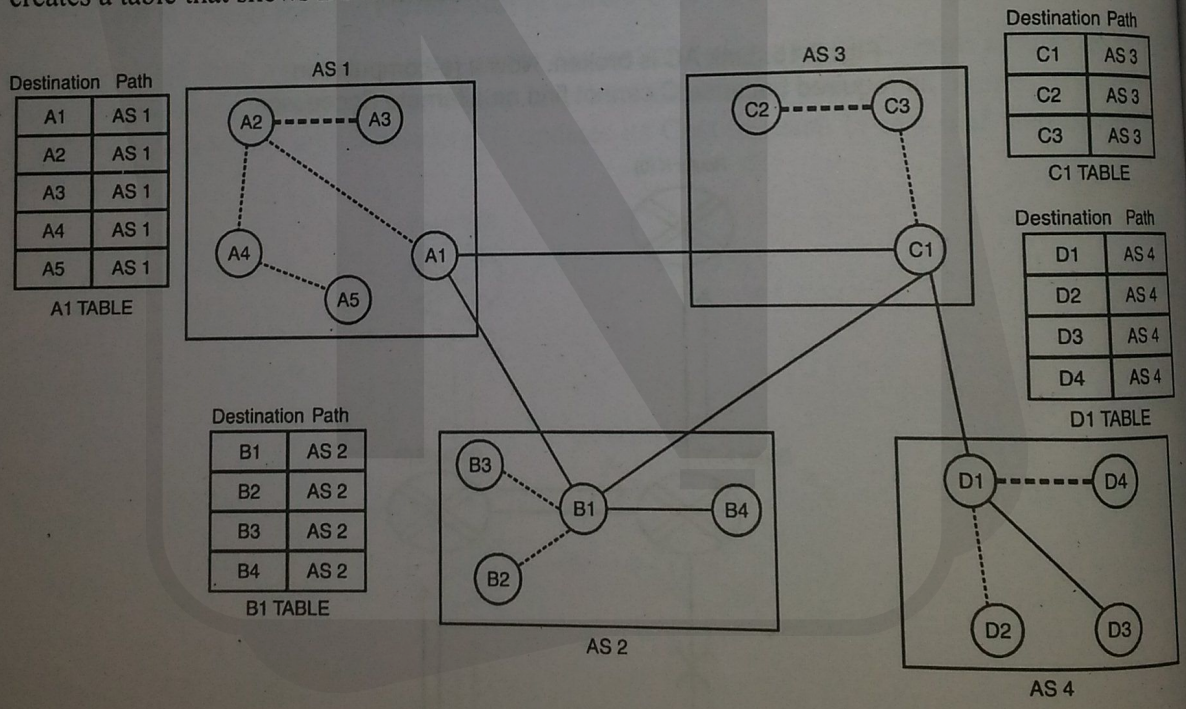

Fig. 2.17.

Phase 2, Advertising: In the second phase, every speaker node advertises its routing table with adjacent speaker nodes. Every speaker advertises the information about different nodes in its AS and their path.

For example, node A1 shares its table with nodes B1 and C1.

A1 shares its table with nodes B1 and C1.

C1 shares its table with nodes D1, B1 and A1.
B1 shares its table with C1 and A1.
D1 shares its table with C1.

When a speaker table node receives a table from a neighbour, it updates its own table by adding the nodes that are not in its routing table. In this way every speaker updates its routing table and adds paths to nodes in other ASs. Example, the routing table of A1 after exchange of these messages looks like:

A1 Table

Destination	Path
A1 ..... A5	AS1 AS1
B1 ..... B4	AS1-AS2 ..... AS1-AS2
C1 ..... C3	AS1-AS3 ..... AS1-AS3
D1 ..... D4	AS1-AS2-AS4 ..... AS1-AS2-AS4

If router A1 receives a packet for nodes A3, it knows that the path is in AS1. Also if A1 receives a packet for D1, it knows that the packet should go from AS1 to AS2 and then to AS3.

Phase 3, Loop prevention and optimization: This phase is concerned with prevention of loops and optimization. Infinite looping is prevented at the router nodes. If a router discovers that its AS is already included in Routing table of an advertised message, it understands that it is a duplicate. The router simply discards the message and ignores it. This prevents the infinite looping problem. The best route may be selected based on path characteristics such as Number of ASs on a path or the total hops. Other factors may also be included in path prioritization such as congestion, cost, etc.

The exchange of Routing information between different ASs is in the form of BGP sessions. These sessions are established between routers for building efficient paths. These sessions may be internal or external depending whether the exchange of information is within a AS or in between two ASs respectively.

CHAPTER 3

Multicasting in IP Environment

3.1 INTRODUCTION

Unicast, Multicast and Broadcast

Unicasting is the Process of Sending a Message from a Single Source to a Single Destination. It is the Communication Process between a Source-Destination Pair. The process of unicast routing uses one of the many algorithms for routing like OSPF, RIP, etc. **Multicasting** is defined as the Process of Sending a Message to a group containing more than one destinations. It is a one to many communication process, where a single message travels to a group of nodes. The process of Multicast routing is different from unicast routing.

The process of multicasting is different from multiple unicasting. In multiple unicasting, multiple copies of the same packet are transmitted by the Sender, one for each destination. The router redirects each packet towards its destination, as per the path available in its routing table. There is always some time lag between first & the last packet transmission by the source. But the problem of multicasting is different. Here only a single copy of a packet is created by the source & this packet is to be transmitted to all nodes of a group. Here the router builds a shortest path tree. The source becomes the root of the tree and all nodes of the group become its leaves. Now the router creates copies of the packet, one for each different path, as defined by the shortest path tree.

Broadcasting is the Process of sending a Message to all nodes of a network. It is one to all communication. Broadcasting is generally required for applications like Route Discovery in Routing Protocols and in algorithms like ARR, RARP, etc. Only a single copy of packet originates from the source and each router recreates Multiple Packets, one for each path connected to the router. Every router transmits the packet only once and duplicate packets are discarded.

3.2 APPLICATIONS OF MULTICASTING

Multicasting finds its use in diverse applications like:

1. Teleconferencing: When multiple speakers connect to each other and interact with each other. Here message of each speaker is multicasted to all other participants.
2. Virtual Classroom: Line sessions where lectures of a professor are multicasted to different students or different Universities at the same time.

3. Banking and Distributed Databases: Banking and many other applications use destiributed databases. A same message has to be transmitted to multiple locations where the database resides.
4. Business Applications: There can be numerous business applications where the same information has to be transmitted to multiple locations at the same time
5. Group Messaging: We often send important official or personal messages to multiple persons with a single click.

3.3 SHORTEST PATH TREES

For multiple routing, a router do not need shortest path to only one node, but shortest paths to all nodes of a group. The problem is that there can be many different groups and a node can participate in multiple groups. Therefore, for routing a tree has to be constructed for every new group. The shortest path tree consists of the source as its root and all member nodes become leaves of the tree. Consider the example below:

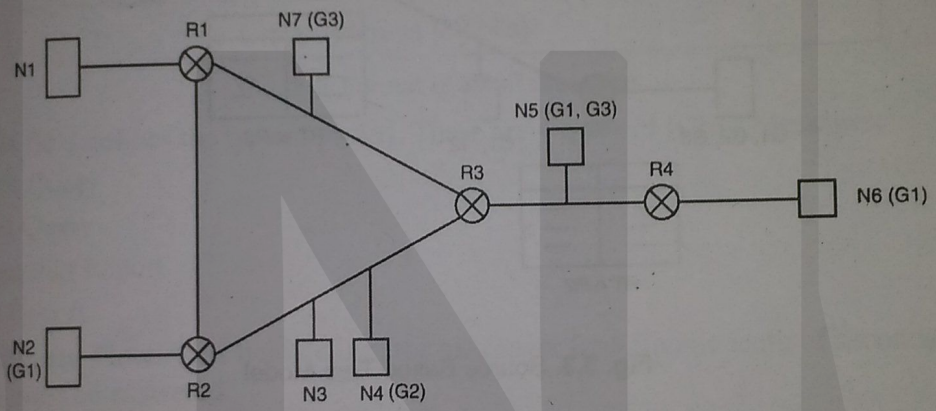

Fig. 3.1. A Sample network with Group Membership

Here we have (G1, G2) routers R1 to R4 and 6 nodes. A node can be a member of more than one group. The group numbers are mentioned against the nodes. G1 consists of Nodes N2, N3, N5 and N6. If N1 needs to send a message to G1, router R1 builds its Routing table as follows:

Group	Next Hop
G1	R3, R2
G2	R2
G3	—, R3

— Routing Table of R1

Fig. 3.2. Routing table for R1 for fig/ 3.1

As we can see in diagram above that a multicast tree is built for every group. To send a packet to G1, R1 must duplicate the packet on R2 and R3. The only problem is the time and resources required to build so many trees because a node can be a member of different group. There are two approaches to store these trees:

1. **Source Based Trees.** In source based tree, every router produces and stores its own Shortest Path Tree (SPT) for every group. The number of SPTs for a router is equal to number of groups.

2. **Group Shared Trees.** In group shared trees, every router does not maintain and store trees for different groups. Rather, a dedicated router is also called **rendezvour router**. This rendezvour router builds SPTs for every group. Whenever a multitasking request arrives at a router, the router simply forwards the request to the rendezvous router . The rendezvous router creates multiple copies of the packet and forwards it on every link as per its group routing table.

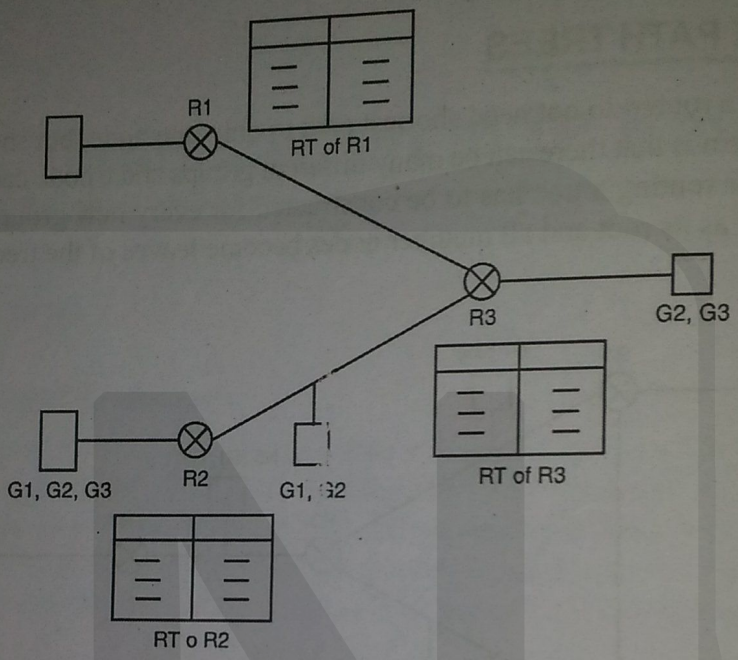

Fig. 3.3. Source Based Tree Model

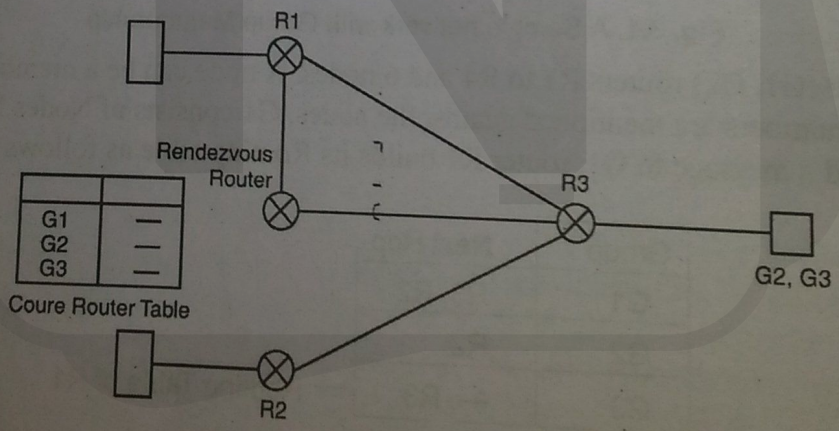

Fig. 3.4. Group Shared Tree Model

3.4 MULTICAST GROUP MEMBERSHIP DISCOVERY PROTOCOLS

Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) are the prominent Multicast Group Membership Discovery protocols. They are essentially the same protocol, with IGMP used for IPv4 multicast groups and MLD used for IPv6 multicast groups. These protocols specifically discover which multicast addresses are of interest to the neighbouring nodes of a

Rendezvous Router. The Rendezvous Router sends Query message from time to time, to discover the membership status of different hosts. Accordingly, this information is utilized in assisting the underlying multicasting routing protocol to make routing decisions. Packets are not broadcasted, but they are sent only over those links, which have active members.

3.4.1 Internet Management Routing Protocol (IGMP)

Multicasting is often needed in Computer Networks. To increase the efficiency of routing protocols, it is necessary to create multicast trees efficiently. We should send multicast packets only to those hosts and routers which have some active members from the group. IGMP is a protocol to manage the group membership of hosts and routers. Every network may have one or more Centre Point Router (Rendezvous Router). These Centre Points are responsible for creation of shortest path trees. All multicast packets are unicasted to Centre Points, which forward them on efficiently. The creation of trees for each group is done with help of IGMP protocol. The format of an IGMP Message is:

Type (8 bits)	Max. Response Time (8 bits)	Checksum (16 bits)
Group Address (32 bits)		

Fig. 3.5. Format of IGMP message

Type: This field defines the type of packet. These are 4 types of IGMP Messages:

i. General Query
ii. Special Query
iii. Membership Report
iv. Leave Report

Max. Response Time: This field defines the maximum time (in one tenth of seconds) in which a query message must be answered.

Checksum: It is a 16 bit checksum for error detection.

Group Address: This field is set to the group address in case of special query, membership report and leave report messages. In case of general query, its value is set to 0.

Membership Report Message: When a host wants to join a group, it sends a Membership Report Message to the Centre Point. The group address of the group of interest is inserted in this message and it is unicasted to the Centre Point which records this membership. This membership report is sent twice to handle transmission losses.

Leave Report Message: When a host is no longer interested in a group, it can send a leave Report Message to the Centre Point. The sending host enters the group address in this message and unicasts it to the Centre Point. Similarly when a router comes to know that none of its networks is interested in a group, it can also send a Lever Report Message to the Centre Point. However, a router does not remove the group immediately from its list because it is possible that some other host or router is still interested in that group. To verify it, this router sends a special query message.

Special Query Message: When a router receives a Leave Report from some of its members, it wants to be sure that some other hosts or router is not wing this group. For this it sends a Special Query Message to that router/host. If that host is still interested in the group, it sends back a Membership Report Message. However if no such message is received within a time frame, then this router removes this group from its list.

General Query Message: This Message is used to Monitor the group membership status over time usually the General Query Message is sent periodically by the Centre Point. The address field in this query message is set to 0.0.0.0. All the hosts and routers which have membership to any group are expected to respond. However, this response is not immediate. Rather, each host/router waits for a random time before responding. This is known as delayed response. During this waiting period, if some other host/router sends its Membership Report for the same group, then it does not send its own membership response. This is done to save sending unnecessary duplicate messages.

3.4.2 Multicast Listener Discovery (MLD) Protocol

The MLD Protocol is similar to the IGMP Protocol. It has the same responsibility as IGMP, i.e. to decide and discover group membership among different hosts of a network. The difference between IGMP and MLD is that while IGMP is used in IPv4, MLD is used in the IPv6. MLD enables routers to discover multicast listeners attached to them. If none of its hosts is an active member of a group, the router conveys this information to the next router on way to the Centre Point (Rendezvous Router). In this case, the upstream router will not forward multicast packets belonging to that group to this router. One or more special routers called the Centre Point hands creation of multicast trees. These multicast trees are created after collecting information from the MLD Protocol. A multicast packet is sent to only those links/routers, which have some active members, as per the information provided by MLD. The MLD header has the following format:

Type (8 bits)	Code (8 bits)	Checksum (16 bits)
Max. Response Delay (16 bits)		Reserved (16 bits)
Multicast Addr. (128 bits)		

Fig. 3.6. Format for MLD Packets

Type: This field identifies the type of packet format. Different type of MLD packets are:

i. Listener Query
ii. Listener Report
iii. Listener Done

Code: This is a 8 bit field to further identify/qualify a packet.

Checksum: This is a 16 bit field. The checksum is a one's complenent of the sum of entire MLD Message alongwith a pseudo-header of IPv6. The pseudo-header contains some basic information of the IPv6 packet.

Max Response Delay: This 16 bits field specifies the maximum time, the sender waits for receiving replies from different hosts/routers in a query message. This time is specified in milliseconds.

Reserved: It is a 16 bit field reserved for future use. It is cleared to 0 by the sender and ignored by all receivers.

Multicast Addr.: This is a 128 bit address of the multicast group. When sending a general query, it is set to 0.

The role of different types of MLD packets are as follows:

Multicast Listener Query: These message are sent by multicast routers in a Querier State to enquire group states of neighboring hosts. When this message is sent, all routers which have some active members of a group respond. These Listener Query Messages can be of 2 types: General or Multicast Addr. Special.

Multicast Listener Reports: These messages are sent by hosts to report to their neighbour routers abut the current group status or any change in the group membership status.

Multicast Listener Done: The Multicast Listener Done message is sent by a host to a multicast router, to signal that there may not by any further group members in the local subnet. When this multicast router receives the Done Message, it verifies the membership of other hosts and routers before deleting it from its table. For this it sends a Listener Query Message.

3.5 MULTICAST ROUTING PROTOCOLS

3.5.1 Multicast Link State Routing (MLSR) Protocol

MLSR is an extension of unicast LCR Protocol. This protocol uses the source based tree method.

Here, each router records the membership of its neighbour nodes to different groups. This membership information is recorded and transmitted to all other routers in form of Link State Advertisement packets (LSAs). When other routers receive all LSAs, the use this information to create a Network topology map for each different group. If there are M number of groups, then m different topologies are built. Then Dijkstra's algorithm is applied on each topology to build M different SPTs. Each SPT services a single group only. The limitation of this protocol is the processing and storage required to build these M diff. SPTs.

3.5.2 Multicast Open Shortest Path Protocol (MOSPF)

It is a direct extension of the OSPF Protocol to the multicast environment. It is a multicast link state routing protocol. Every router determines the groups associated with nods attached to it. This information is stored in a different LSA Packet called 'Group Membership LSA'. This Group Membership LSA is propagated to other routers. In this way every router builds information about all members of a particular group. Dijksta's algorithm is applied to build shortest path to each member of a group. Unicast address of a node is used for calculation purposes in Dijksta's algorithm. All shortest paths to members of a particular group are combined to create a SPT. These SPTs are stored in Cache for future use. To increase the efficiency, these SPTs are built and are not computed and stored automatically. They are computed only when the first request for multicasting to that group arises.

3.5.3 The Distance Vector Multicast Routing Protocol (DVMRP)

Though DVMPR is an extension of the DVR protocol to multicasting environment, it is quite different from the unicast DVR. But inherating the basic characteristics of DVR i.e. a router shares information only, with its neighbour routers. While forwarding a packet, following considerating must be kept in mind:

a. Loops must be prevented
b. Duplicate packets should not be transmitted
c. Membership information of a node must be dynamically updated
d. Shortest paths must be explored.

To fulfill above objectives, DVMPR uses one of the following four approaches, where each one is an improvement over the earlier:

1. **Flooding :** In this strategy, a router simply forwards an arriving packet to all links except the one on which it arrived. The router does not even consult its table. The problem with this starategy is that duplicate packets are created unnecessarily which may cause looping. Looping can be controlled to some extent by keeping a cached copy of forwarded packet for some time and always checking for duplicates when a new packet arrives.
   This will prevent duplicates, but this strategy is not efficient as it creates lot of unnecessary traffic.
2. **Reverse Path Forwarding (RPF) :** Here multiple copies of same packet are prevented, whenever a packet arrives at a router, the router verifies if it has travelled the shortest path while reaching it or not. If the packet has travelled the shortest path to it, only then the packet is forwarded to other links, otherwise it is discarded. The shortest path is verified by the router by looking at its routing table for the address of the source. If the node from which the packet has arrived is the 'Next Node' in its routing table, it means that packet has travelled the shortest path and it is forwarded. Otherwise, the packet is discarded. This strategy prevents loops as the shortest path is always one.
3. **Reverse Path Broadcasting (RPB) :** While RPF prevents looping, but multiple duplicate packets can still reach a node. This is because after checking the shortest path RPF simply forwards a parent on all links. To prevent this, a packet router is designated for each network. After checking the shortest path, a router now sends the packet to only that line for which it is the parent. This way flooding is prevented and duplicate packets do not arrive because a network can receive packets from one and any one router.
4. **Reverse Path Multicasting (RPM) :** In all previous schemes, multicasting packets reach even those networks which have no associated members of the group. We should try to limit transmission to networks where there are no active members of the group. This is done in RPM where there exist a parent router maintaining group membership information of all its nodes. This information is gathered by the parent router from 1GMP protocol. This protocol works by defining 2 operations:
   a. **Pruning:** If none of the nodes of a router is a member of a group, this parent router sends a prune message to its upstream router. The upstream router in turn records this message to this router. Also, a router which receives prune message from all its downstream routers, sends a prune message to its upstream router. Thus unwarranted message to non-member groups are never sent.
   b. **Grafting:** If a router has already sent a prune message to its upstream router, but later it disscuess that one of its nodes, is now associated with that group (due to new information received from (IGMP), then this router sends a graft message to its upstream router. The upstream router will now again start sending messages of that group to this router.

Thus using pruning and grafting, one can easily control sending unnecessary packets to non-member network. This increases the efficiency of RPM Scheme of DVMRP Protocol.

3.5.4 Core Based Tree Protocol (CBT)

In CBT Protocol, each Autonomous System (AS) is divided into regions and we elect one router from each region as the core router. A tree is created for transmission to each group, where the core becomes

the root of the tree. Initially all routers which want to join a group, send a Unicast join message to the core. All intermediate nodes forward this message to the core. When all these join messages reach the core, the core router constructs a tree to reach all member routers. Now, the multicast tree is constructed. Any node which wants to multicast to the group, now unicasts the message to the core. The core router then forwards that message to all member routers of the tree.

3.5.5 Protocol Independent Multicast (PIM)

PIM refers to two different protocols: Protocol Independent Multicast-Dense Mode (PIM-DM) and Protocol Independent Multicast-Sparse Mode (PIM-SM)

1. **Protocol Independent Multicast-Dense Mode:** The PIM-DM is used when most of the routers are a part of the multicasting tree. The working of PIM-DM is like DVMRP. It uses RPM strategy of DVMRP along with pruning and grafting to construct the multicast tree. Whenever a group packet arrives at a router, the router first checks if it has travelled the shortest path while reaching it. If it has, only then the packet is considered for forwarding otherwise not. Also, every network has a parent router, which maintains group information of all its members. A packet is multicasted by a Parent-router if:
   a. The packet has travelled the shortest path in reaching it.
   b. The packet is forwarded only to those links where some active members of the group exist. If none of its members is an active member of the group, the packet is not Multicasted. Grafting Message can be used to resume Sending Messages to a node, which was earlier a non-member. The difference between PIM-DM and DVMRP is that PIM-DM does not use a specific unicast routing protocol unlike DVMRP. It can use either Routing Information Protocol (RIP) or Open Shortest Path First (OSPF) for unicasting.
2. **Protocol Independent Multicast-Sparse Mode (PIM-SM) :** This Multicasting Method can be used when the network operates in Spasse Mode i.e. not all routers participate in a group. In this case, a group-shared tree approach is more appropriate. The working of PIM-SM is similar to the Core-Based Tree (CBT) Method. One of the router serves as Centre Point, which stores the multicast tree. All nodes desirous of multicasting, The Centre Point forwards the packet along the links, as defined by its multicast tree. This tree is formed form the membership information sent by each node to the Centre Point. PIM-SM is more resilient to failures, as it provides backup Centre Point, if there is a failure.

   One unique feature of PIM-SM is that if the network operates in Dense Mode, then instead of sharing one Centre Point between all routers of network, it switches to source based tree approach i.e. every node maintains its own tree information.

■■■