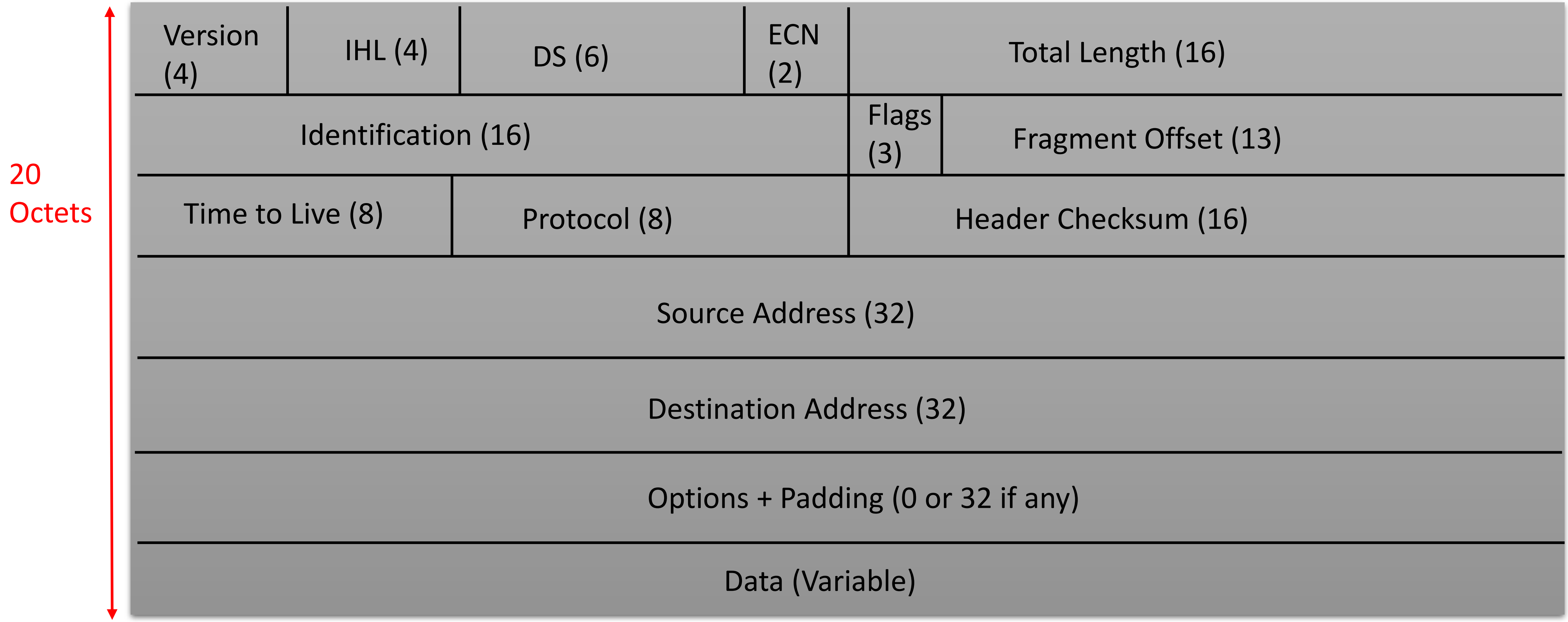
The Internet Protocol (IP)

- The Internet Protocol (IP) is the core of the TCP/IP protocol suite and its main protocol at the network layer.
- IP has four basic functions:
- Addressing - In order to deliver datagrams, IP must know where to deliver them to. For this reason, IP includes a mechanism for host addressing.
- Data Encapsulation - IP accepts data from the transport layer protocols(UDP and TCP). It then encapsulates this data into an IP datagram using a special format prior to transmission.
- Fragmentation and Reassembly - IP datagrams are passed down to the data link layer for transmission on the local network. However, the maximum frame size of each data link network may be different. For this reason, IP includes the ability to fragment IP datagrams into pieces so they can each be carried on the local network. The receiving device uses the reassembly function to recreate the whole IP datagram again.
- Routing - When an IP datagram is sent to a destination on the same local network, this can be done easily using the network's underlying LAN protocol. However, when the destination is on a distant network not directly attached to the source, the datagram must be delivered by routing the datagram through intermediate devices (called routers). IP accomplishes this with support routing protocols.

IPv4 Packet Format

- Data transmitted over an internet using IP is carried in messages called IP datagrams.
- IP datagram consists of following fields:



20 Octets

<table>
<tr><td colspan="4">Version (4)</td><td colspan="4">IHL (4)</td><td colspan="6">DS (6)</td><td colspan="2">ECN (2)</td><td colspan="16">Total Length (16)</td></tr>
<tr><td colspan="16">Identification (16)</td><td colspan="3">Flags (3)</td><td colspan="13">Fragment Offset (13)</td></tr>
<tr><td colspan="8">Time to Live (8)</td><td colspan="8">Protocol (8)</td><td colspan="16">Header Checksum (16)</td></tr>
<tr><td colspan="32">Source Address (32)</td></tr>
<tr><td colspan="32">Destination Address (32)</td></tr>
<tr><td colspan="32">Options + Padding (0 or 32 if any)</td></tr>
<tr><td colspan="32">Data (Variable)</td></tr>
</table>

IPv4 Packet Format

- Version (4 bits) - Indicates version number, the value is 4. By looking at the version number, the router can determine how to interpret the remainder of the IP datagram.
- Internet Header Length (IHL) (4 bits) - Because an IPv4 datagram can contain a variable number of options these 4 bits are needed to determine where in the IP datagram the data begins. Most IP datagrams do not contain options, so the typical IP datagram has a 20 byte header.
- Differentiated Service(DS)/Explicit Congestion Notification(ECN)(8 bits) - Allows to mark packets for differentiated treatment to achieve Quality-Of-Service (QoS), e.g. express priorities. The ECN field provides for explicit signaling of congestion.
- Total Length (16 bits) - Total datagram length, including header plus data, in bytes. Since this field is 16 bits long, the theoretical maximum size of the IP datagram is 65,535 bytes.
- Identification (16 bits) - A sequence number that, together with the source address, destination address, and user protocol, is intended to identify a datagram uniquely. Thus, this number should be unique for the datagram's source address, destination address, and user protocol for the time during which the datagram will remain in the internet.
- Flags (3 bits) - Only two of the bits are currently defined. MF is 'more fragments' and is used for fragmentation and reassembly. The DF 'Don't Fragment' bit prohibits fragmentation when set.
- Fragment Offset (13 bits) - When fragmentation of a message occurs, this field specifies the offset, or position, in the overall message where the data in this fragment goes. It is specified in units of 8 bytes (64 bits).

IPv4 Packet Format

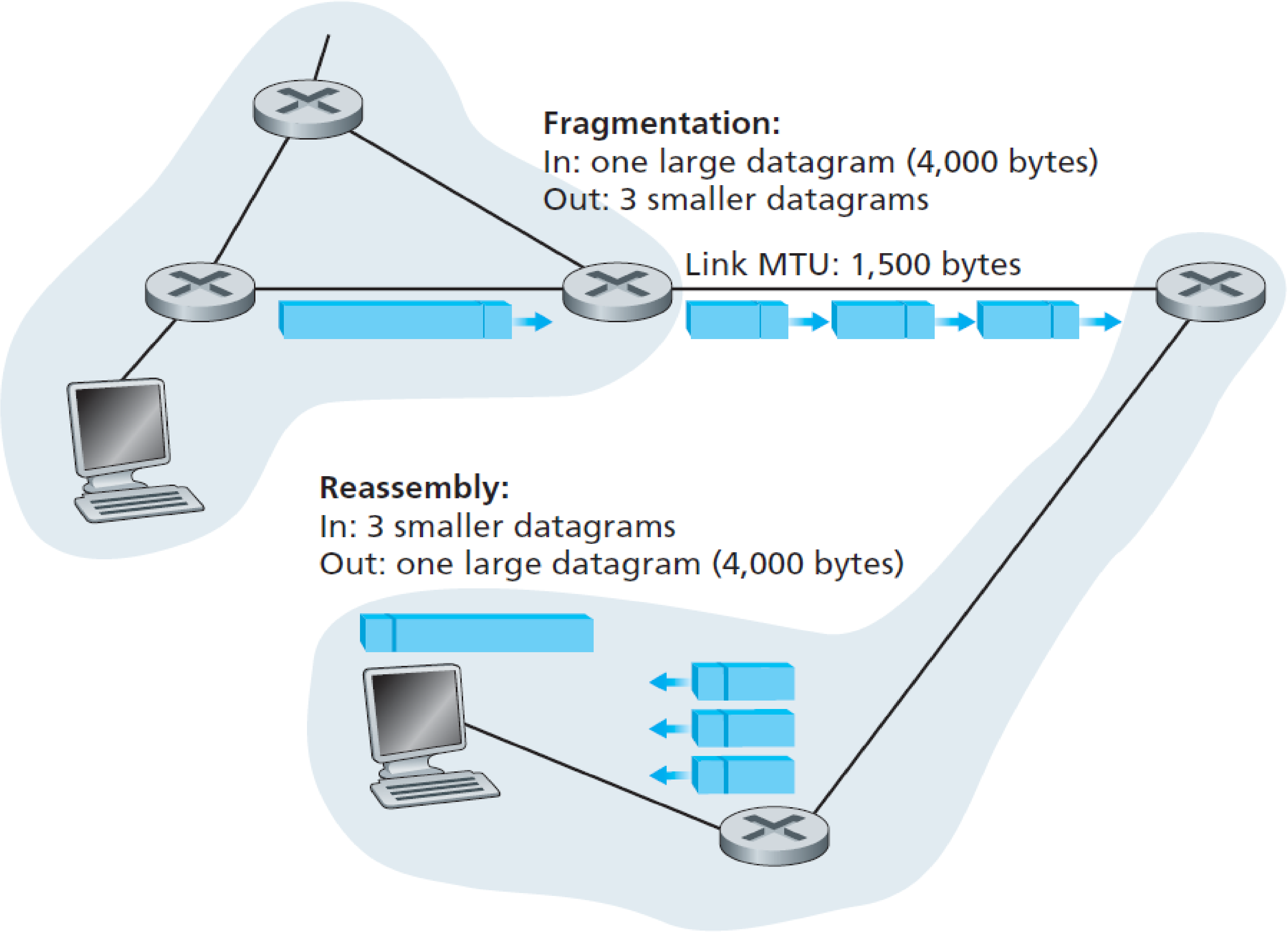
- Time to Live (8 bits) - The time to live (TTL) field is included to ensure that datagrams do not circulate forever in the network. It specifies how long the datagram is allowed to 'live' on the network, in terms of router hops .This field is decremented by one each time the datagram is processed by a router. If the TTL field reaches 0, the datagram must be dropped.
- Protocol (8 bits) - Indicates the next higher level protocol(either transport layer protocol or encapsulated network layer protocol) that is to receive the data field at the destination. Example values are ICMP = 0x01 , TCP = 0x06 , UDP = 0x11.
- Header Checksum (16 bits) - An error detecting code applied to the header only. The header checksum is computed by treating each 2 bytes in the header as a number and adding these numbers using 1s complement arithmetic. The 1s complement of this sum, known as the Internet checksum, is stored in the checksum field. A router computes the header checksum for each received IP datagram and detects an error condition if the checksum carried in the datagram header does not equal the computed checksum. Routers typically discard datagrams for which an error has been detected. The checksum must be recomputed and stored again at each router, as the TTL field, and possibly the options field as well, may change.
- Source and Destination Addresses (32 bits each) - When a source creates a datagram, it inserts its IP address into the source IP address field and inserts the address of the ultimate destination into the destination IP address field.
- Options/Padding (variable) - Contains header field for optional IP feature requested by the sending user. If one or more options are included, and the number of bits used for them is not a multiple of 32, enough zero bits are added to make the header to a multiple of 32 bits (4 bytes).
- Data/Payload (variable) - The data field of the IP datagram contains the transport layer segment (TCP or UDP) to be delivered to the destination. However, the data field can carry other types of data, such as ICMP messages. The data field must be an integer multiple of 8 bits in length. The maximum length of the datagram (data field plus header) is 65,535 octets.

IP Fragmentation & Reassembly

- The maximum amount of data that a link layer frame can carry is called the maximum transmission unit (MTU). For example, Ethernet frame can carry up to 1500 bytes of data.
- Because each IP datagram is encapsulated within the link layer frame for transport from one router to the next router, the MTU of the link layer protocol places a hard limit on the length of an IP datagram. Another issue is that each of the links along the route between sender and destination can use different link layer protocols, and each of these protocols can have different MTUs.
- The solution is to fragment the data in the IP datagram into two or more smaller IP datagrams, encapsulate each of these smaller IP datagrams in a separate link layer frame; and send these frames over the outgoing link. Each of these smaller datagrams is referred to as a fragment.
- In addition, every intermediate router can either fragment a full message or further fragment a fragment when necessary for transmission on next hop.
- Fragments need to be reassembled before they reach the transport layer at the destination. The question is : Where they should be reassembled? Network Routers or Destination End System!
- Reassembly at intermediate routers can have following disadvantages:
  - Large buffers are required at routers, and there is the risk that all of the buffer space will be used up storing partial datagrams.
  - All fragments of a datagram must pass through the same router which can prevent the use of dynamic routing.
- Thus, datagram fragments are reassembled at the destination end system.

IP Fragmentation & Reassembly

- All fragment datagrams belonging to same message have:
  - A full IP header
  - Identification field(ID) – same for all fragments.
  - Total Length field - reflecting the fragment size.
  - Fragment Offset field – different for all fragments, reflecting the start of the present fragment within the whole message, specifies offset in multiples of 64 bits.
  - MF Flag (more fragments) bit – set for all fragments except for the last fragment.
- When a datagram is created, the sending host stamps the datagram with an identification number, source and destination addresses. When a router needs to fragment a datagram, each resulting datagram (that is, fragment) is stamped with the source address, destination address, and identification number of the original datagram.
- Let’s assume, datagram of 4000 bytes wide (including the 20 byte IP header) needs to be sent over a link with an MTU of 1500 bytes. Suppose original datagram has an identification number of 123. The following steps are taken at the router:
- Create First Fragment – Total Length(Bytes) = 1480 , ID = 123 , Fragment offset = 0 (data should be inserted beginning at byte 0) , MF = 1 (more fragments also).
- Create Second Fragment – Total Length(Bytes) = 1480 , ID = 123 , Fragment offset = 185 (data should be inserted beginning at byte1480) , MF = 1 (more fragments also).
- Create Third Fragment – Total Length(Bytes) = 1020(3980-1480-1480) , ID = 123 , Fragment offset = 370 (data should be inserted beginning at byte 2960) , MF = 0 (this is the last fragment).



Fragmentation:
In: one large datagram (4,000 bytes)
Out: 3 smaller datagrams

Reassembly:
In: 3 smaller datagrams
Out: one large datagram (4,000 bytes)

IP Fragmentation & Reassembly

- To reassemble a datagram following steps are taken:
- The receiving device initializes a buffer where it can store the fragments of the message as they are received.
- The receiving device sets up a timer for reassembly of the message.
- As fragments with the same ID arrive, their data fields are inserted in the proper position in the buffer until the entire data field is reassembled, which is achieved when a contiguous set of data exists starting with an Offset of zero and ending with data from a fragment with a false More Flag.
- The IP service does not guarantee delivery. If the timer for the reassembly expires with any of the fragments missing, the message cannot be reconstructed. The already received fragments are discarded, and an ICMP message is generated for the source host.
- Fragmentation/Reassembly creates significant overhead:
    - Several datagrams transmitted per message, each one having full IP header.
    - Complicates router and end systems which need to be designed to accommodate fragmentation/reassembly.
    - Upon loss of single fragment the whole message is possibly retransmitted by higher layers.

IP Routing

- IP routing is the process of moving packets from one network to another network using routers.
- To accomplish this, a path or route through the network must be determined.
- It is possible the more than one route is also available. Thus, a routing function must be performed.
- Following requirements are imposed on the routing function:
  - Correctness - computed routes should be valid paths that contain no loops.
  - Simplicity - routing algorithms / protocols should be computationally simple and require only little information exchange among routers.
  - Robustness - a routing protocol must be able to cope with: link or station failures, newly established links or stations, changes in link metrics , congestion situations by establishing new routes when old ones become infeasible or are no longer optimal.
  - Stability - a routing protocol should not recompute everything upon minor changes in the network.
  - Fairness – all users should be treated in equal manner.
  - Optimality – It has different perceptive based on different criteria. From user perspective, generated routes should be short, fast and offer good throughout. From provider perspective, the network should carry as many packets as possible.
  - Efficiency - when a route between two nodes exists, the routing algorithm / protocol should be able to find it.

IP Routing – Performance Criteria

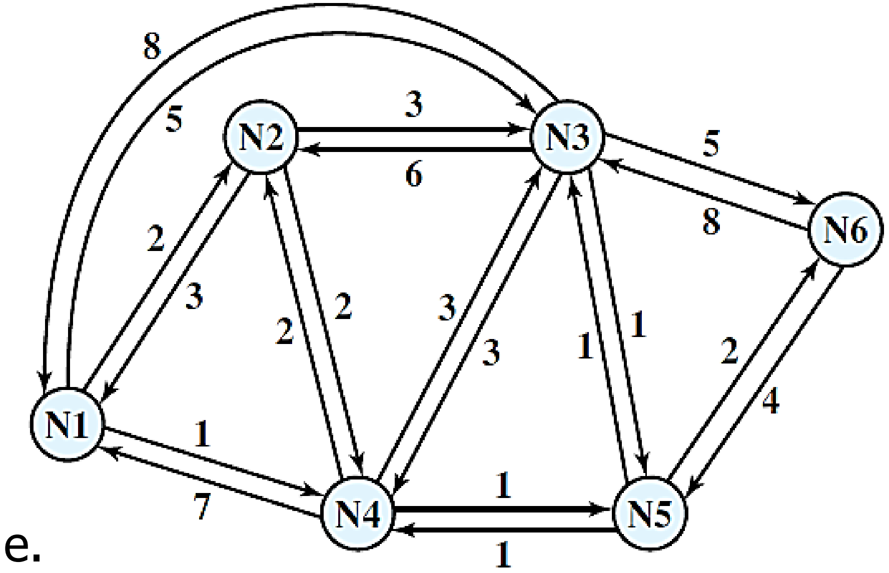
- The selection of a route is based on following criterion:
- Number of Hops or least cost Criteria- Choose the minimum hop route (one that passes through the least number of nodes) through the network. In least cost routing, a cost is associated with each link, and, for any pair of attached stations, the route through the network that accumulates the least cost is chosen.
  - Shortest path(fewest hops) from N1 to N6:

    Nodes visited $= N1 \rightarrow N3 \rightarrow N6$

    Cost $= 5 + 5 = 10$
  - Least Cost path from N1 to N6:

    Nodes visited $= N1 \rightarrow N4 \rightarrow N5 \rightarrow N6$

    Cost $= 1 + 1 + 2 = 4$



- Decision Time – refers to when routing the decisions are made i.e.

for individual packet or for each session or at the time of network configuration time.

IP Routing – Performance Criteria

- Decision Place - refers to which node or nodes in the network are responsible for the routing decision. In distributed routing, each node has the responsibility of selecting an output link for routing packets as they arrive. In centralized routing , the decision is made by some designated node, such as a network control center. In source routing the routing decision is made by the source station rather than by a network node and is then communicated to the network.
- Network Information Source – refers to the information used for making routing decision such as knowledge of the topology of the network, traffic load, and link cost. Some strategies use no such information and manage to get packets through flooding. In distributed routing, the individual node may make use of only local information, such as the cost of each outgoing link. Each node might also collect information from adjacent (directly connected) nodes, such as the amount of congestion experienced at that node. In centralized routing, the central node typically makes use of information obtained from all nodes.
- Information update timing – refers to when information used in routing decision is updated i.e. the information is never updated or it is updated periodically to enable the routing decision to adapt to changing conditions. Thus the more frequently it is updated, the more likely the network is to make good routing decisions.

Routing Strategies

- Four key strategies are:
- Fixed Routing - A single, permanent route is configured for each source destination pair of nodes in the network. Either of the least cost routing algorithms can be used. The routes are fixed, or at least only change when there is a change in the topology of the network. The advantage of fixed routing is its simplicity, and it should work well in a reliable network with a stable load. Its disadvantage is its lack of flexibility. It does not react to network congestion or failures.
- Flooding - This technique requires no network information. A packet is sent by a source node to every one of its neighbors. At each node, an incoming packet is retransmitted on all outgoing links except for the link on which it arrived. Flooding technique is highly robust and could be used to send emergency messages. The principal disadvantage of flooding is the high traffic load that it generates, which is directly proportional to the connectivity of the network.
- Random Routing - With random routing, a node selects only one outgoing path for retransmission of an incoming packet. The outgoing link is chosen at random, excluding the link on which the packet arrived. If all links are equally likely to be chosen, then a node may simply utilize outgoing links in a round robin fashion. A probability can also be assigned to each outgoing link and link is selected based on that probability. Like flooding, random routing requires the use of no network information.
- Adaptive Routing - The routing decisions are made change as conditions on the network change. The principal conditions that influence routing decisions are node or link failure and congestion. For adaptive routing to be possible, information about the state of the network must be exchanged among the nodes.

Routing Algorithm & Routing Protocols

- A network is modelled as a graph $G = (N, E)$ which is a set N of nodes and a collection E of edges. In the context of network layer routing, the nodes in the graph represent routers and the edges connecting these nodes represent the physical links between these routers.
- A host is directly attached to one router called as the default router or the first hop router. Whenever a host sends a packet, the packet is transferred to its default router. The default router of the source host is referred as the source router and the default router of the destination host is referred as the destination router.
- The routers in an internet are responsible for receiving and forwarding packets through the interconnected set of networks. Each router makes routing decision based on knowledge of the topology and traffic/delay conditions of the network.
- The purpose of a routing algorithm is simple: given a set of routers, with links connecting the routers, a routing algorithm finds a good path from source router to destination router.
- A routing protocol specifies how routers communicate with each other to distribute routing information that enables them to select routes between two nodes.

Routing Tables

- Each router maintains a set of information that provides a mapping between different network IDs and the other routers to which it is connected. This information is contained in a routing table.
- Each entry in the table is called a routing entry which provides information about one network.
- Each time a packet is received, the router checks its destination IP address against the routing entries in its table to decide where to send the packet, and then sends it on its next hop.
- The routing table contains information not only about the networks directly connected to the router, but also information that the router has learned about more distant networks.
- Common fields in routing table are:
- Destination IP address – it can be a host address or network address to which the packet is finally delivered.
- Next hop address – it is the address of the next hop router to which the packet is delivered.
- Outgoing Interface - used when forwarding the packet to the next hop or final destination.
- Flags - A flag telling whether destination IP is host or network; A flag telling whether next hop is a router or directly attached network
- A routing table can be static or dynamic. Static routing table contains information that is entered manually. The administrator enters the route for each destination into the table. When a table is created, it cannot update automatically when there is a change in the network. The table must be manually altered by the administrator. A dynamic routing table is updated periodically by using one of the dynamic routing protocols whenever there is a change in the network.