UNIT 4 INTERNETWORKING

Structure **Page Nos.**

4.0 Introduction 52
4.1 Objectives 52
4.2 Internetworking 52
4.2.1 How does a Network differ?
4.2.2 Networks Connecting Mechanisms
4.2.3 Tunneling and Encapsulation
4.3 Network Layer Protocols 55
4.3.1 IP Datagram Formats
4.3.2 Internet Control Message Protocol (ICMP)
4.3.3 OSPF: The Interior Gateway Routing Protocol
4.3.4 BGP: The Exterior Gateway Routing Protocol
4.4 Summary 68
4.5 Solutions/Answers 68
4.6 Further Readings 69

4.0 INTRODUCTION

There are many ways in which one network differs from another. Some of the parameters in which a network differs from another network are packet length, quality of services, error handling mechanisms, flow control mechanism, congestion control mechanism, security issues and addressing mechanisms. Therefore, problems are bound to occur when we require interconnection between two different networks. Different mechanisms have been proposed to solve this problem: Tunneling is used when the source and destination are the same type of network but, there is a different network in-between. Fragmentation may be used for different maximum packet sizes of different networks. The network layer has a large set of protocols besides IP. Some of these are OSPF and BGP and ICMP. In this unit, we will discuss some of these protocols as well as some internetworking mechanisms.

4.1 OBJECTIVES

After going through this unit, you should be able to:

- list how are network differs from another;
- list the components of a network layer;
- define tunneling and fragmentation concepts;
- discuss the field format of IP datagram, IP addressing;
- describe internet routing protocols, such OSPF and BGP, and
- introduce Internet Control Message Protocol (ICMP).

4.2 INTERNETWORKING

The Internet is comprised of several networks, each one with different protocols. There are many reasons for having different networks (thus different protocols):

- Many personal computers use TCP/IP.
- Many larger business organisations still use IBM mainframes with SNA Protocol.

- Some PCs still run on Novell's NCP/IPX or Appletalk.
- Wireless Network will have different protocols.
- A large number of Telecommunication companies provide ATM facilities.

In this section, we will examine some issues that arise when two or more networks are interconnected. The purpose of interconnecting is to allow any node or any network (e.g., Ethernet) to access data to any other node on any other network. (e.g., ATM). Users should not be aware of the existence of multiple networks.

4.2.1 How does a Network differ?

Tanenbaum [*Ref.1*] has defined several features (Kindly refer to *Table1*) that distinguishes one network from another. These differences have be resolved while internetworking. All these features are defined at the network layer only, although, the network differs at the other layers too. They might have different encoding techniques at the physical layer, different frame formats at the data link layer, and different Q_0 S at the transport layer etc.

Table 1: Different type of Networks

Features	Options
Types of services	Connection-oriented, connection-less,
Protocols	IP, IPX, SNA, ATM
Addressing Scheme	Flat vs. Hierarchical
Maximum Packet size	Different for each network
Flow Control	Sliding window, Credit based
Congestion Control Mechanism	Leaky bucket, Token bucket, Hop by Hop, Choke Packets
Accounting	By connect time, packet by packet, byte by byte

4.2.2 Networks Connecting Mechanisms

We have been addressing the problems of connecting network in the earlier blocks also. Let us revisit these topics again. Networks can be connected by the following devices:

- **Repeaters or Hubs** can be used to connect networks at the physical layer. The purpose of these devices is to move bits from one network to another network. They are mainly analog devices and do not understand higher layer protocols.
- At data link layer, bridges and switches were introduced to connect multiple LANs. They work at the frame level rather than bits level. They examine the MAC address of frames and forward the frames to different LANs. They may do little translation from one protocol (e.g., Token ring, Ethernet) to another MAC Layer Protocol. Routers were used at the network layer, which also does translation incase the network uses different network layer protocols.

Finally, the transport layer and application layer gateways deal with conversion of protocols at the transport and application layer respectively, in order to interconnect networks.

The focus of the section is to introduce mechanism internetworking at the network layer. Therefore, we have to understand the difference between the **switching that** operates at the data link layer and routing that operate at the network layer.

The main difference between the two operations is that, with a switch, the entire frame is forwarded to a different LAN on the basis of its MAC address. With a router, the packet is extracted and encapsulated in a different kind of a frame and forwarded to a remote router on the basis of the IP address in the packet. Switches need not understand the network layer protocol to switch a packet, whereas, a router requires to do so. **Tanenbaum** [*Ref.1*] has described two basic mechanisms of internetworking: **Concatenated virtual circuit** and **connectionless internetworking**. In the next sections we will talk about them.

Concatenated Virtual Circuit

This scheme is similar to the implementation of a connection-oriented service in which, a connection from the source router to the destination router must be established before any packet can be forwarded. This type of a connection is called a virtual circuit, keeping with the analogy of the physical circuits set up by the telephone system and the subnet is called a Virtual Circuit Subnet. The idea behind a Virtual Circuit is to avoid choosing a new route for every packet to be forwarded. A route selected as a part of the connection setup is stored in tables inside the routers.

The essential feature of this mechanism is that a series of Virtual Circuits is setup from the source machine on one network to the destination machine on another network through one or more gateways. Just like a router, each gateway maintains tables, indicating the virtual circuits that are to be used for packet forwarding. This scheme works when all the networks follow the same Q_0S parameters. But, if only some networks support reliable delivery service, then all the schemes will not work. In summary, this scheme has the same advantage and disadvantage of a Virtual Circuit within a subnet.

Datagram Model

Unlike the previous model there is no concept of virtual circuits, therefore, there is no guarantee of packet delivery. Each packet contains the full source and destination address and are forwarded independently. This strategy uses multiple routes and therefore, achieve higher bandwidth.

A major disadvantage of this approach is that, it can be used over subnets that do not use Virtual Circuit inside. For example, many LAN and some mobile networks support this approach.

In summary, approach to internetworking is the same as datagram subnets: Congestion proves, robustness in case of router failures.

4.2.3 Tunneling and Encapsulation

It is used when the source and destination networks are the same but the network, which lies in-between, is different. It uses a mechanism called encapsulation where, a data transfer unit of one protocol is enclosed inside a different kind of protocol. Tunneling allows us to carry one kind of frame that uses a particular network but uses, a different kind of frame.

Suppose two hosts located very far away from each other wants to communicate and both have access to the Internet link. It means that both of them are running TCP/IP based protocol. The carrier (WAN) which lies between the two hosts is based at X.25. Its format is different from TCP/IP. Therefore, the IP datagram forwarded by the host one will be encapsulated in X.25 network layer packet and will be transported to the address of the router of the destination host, when it gets there. The destination router removes the IP packet and sends it to host 2. WAN can be considered as a **big tunnel**

extending from one router to another [*Ref.1*]. The packet from host 1 travels from one end of a X.25 based tunnel to another end of the tunnel encapsulated properly. Sending and receiving hosts are not concerned about the process. It is done by the concerned router at the other end.

☞ Check Your Progress 1

1) List the important features in which one network differs from another.

……

……

……

2) What are the mechanisms for interconnecting networks?

……

……

……

3) Where is tunneling used?

……

……

……

4.3 NETWORK LAYER PROTOCOLS

In the network layer, the internet can be viewed as a collection of subnetworks or Autonomous systems that are interconnected. End systems are not usually directly attached to each other via a single communication link. Instead, they are directly connected to each other through intermediate switching devices known as routers. A router takes a chunk of information arriving on, one of its incoming links and forwards that chunk of information on one of its outgoing communication channels. Routing and communication links are a part of Internet service providers. To allow communication among Internet users and to allow users to access worldwide internet content, these lower tier ISPs are interconnected through national and international upper tier ISPs. An upper tier ISP consists of high speed routers interconnected with high speed fiber-optic channels [*Ref.5*]. Each ISP network whether upper tier or lower tier, is managed independently and runs the IP Protocol which holds the whole internet together and provides a best-efforts (not guaranteed) service to forward information (datagrams) from source to destination without regards to where the machines are located. In this section, we will introduce several network layer protocols. The following *Figure 1* shows the network layer with these major components [*Ref.5*]:

(i) IP

(ii) ICMP

(iii) RIP, OSPF and BGP

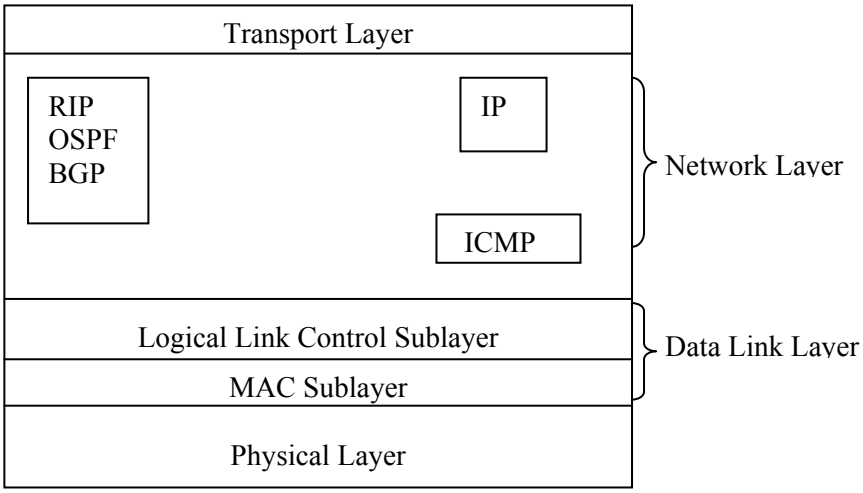

Figure 1: Network layer protocols

(i) **IP:** The first component is IP Protocol, which defines the followings:

- Fields in IP datagram
- Address formats
- Action taken by routers and end systems on a IP datagram based on the values in these fields.

(ii) **ICMP:** The Second Component of the network layer is ICMP (Internet Control Message Protocol) used by the hosts, routers and gateways to communicate network layer information to each other. The most typical use of ICMP is for error reporting.

(iii) **RIP, OSPF and BGP:** The third component is related to routing protocols: RIP and OSPF are used for Intra-AS routing, whereas, BGP is used as exterior gateway routing protocol.

Now we will describe each component separately.

4.3.1 IP Datagram Formats

An IP datagram consists of a header part and a data part. The header has a 20-byte fixed part and a variable length optional part as shown in the *Figure2(a)*. The header format is shown in *Figure 2(b)*. It is transmitted in big-endian order: from left to right, with the high-order bit of the *Version* field going first. On little endian machines, software conversion is required on both the transmission header as well as the reception header. The key fields in the IPv_4 Internet Protocol version 4 datagram header are the followings.

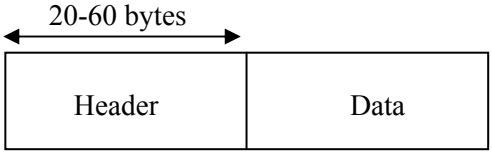

Figure 2 (a) : IP datagram

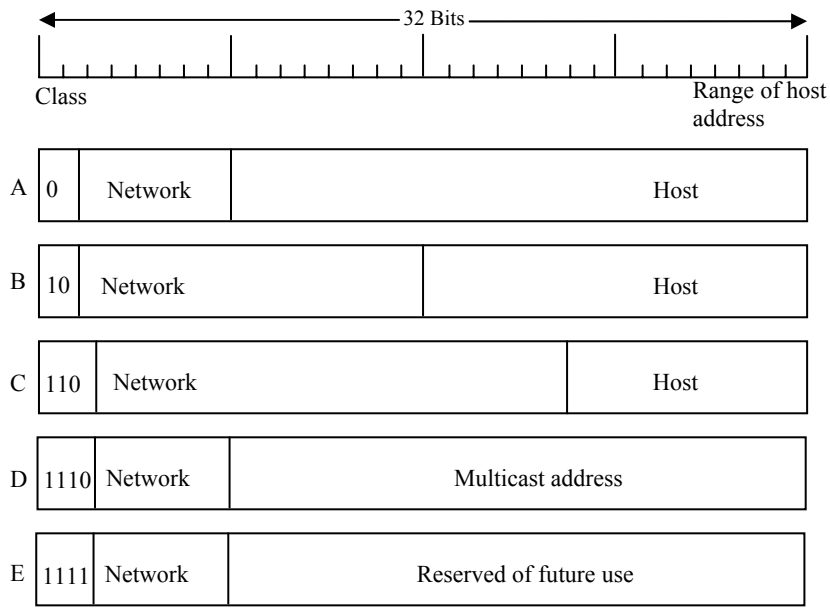

Figure 2 (b): IP address formats

The *Version* field specifies the IP version of the protocol, the datagram belongs to. By including the version in each datagram, the router can determine/interpret the reminder of the IP datagram.

Header length (4 bits)

The field defines the length of the header in multiples of four bytes. The four bytes can represent a number between 0 and 15, which when multiplied by 4, results in a 60 bytes. A typical IP datagram has 20 byte header only, because most IP datagram do not contain options.

The *Type of service* (8 bits) defines how the datagram should be handled. It defines bits to specify priority, reliability, delay, level of throughput to meet the different requirements. For example, for digitised voice, fast delivery, beats accurate delivery. For file transfer, error-free transmission is more important than fast transmission.

The *Total length* includes everything in the datagram both header and data. The maximum length is 65,535 bytes. At present, this upper limit is tolerable, but with future gigabit networks, larger datagrams may be needed.

The *Identification* filed is used in **fragmentation.** A datagram when passing through different networks may be broken into several fragments to match the network frame size. Once the fragmentation occurs, each fragment is identified with a sequence number to this field. Surprisingly, the IPV_6 does not permit the fragmentation of the IP datagram at the routers level.

Next, comes an unused bit and then two 1-bit fields. ***DF*** stands for **Don't Fragment**. It is an order to the router not to fragment the datagram because, the destination is incapable of putting the pieces back together again.

MF stands for **More Fragments**. All fragments except the last one have this bit set. It is needed to know when all fragments of a datagram have arrived.

The ***Fragment offset*** (13 bits) depicts the location of that the current datagram, this fragment belongs to. All fragments except, the last one in a datagram, must be a

multiple of 8 bytes, the elementary fragment unit. Since 13 bits are provided, there is a maximum of 8192 fragments per datagram, with a maximum datagram length of 65,536 bytes, one more than the *Total length field.*

The *Time to live* (8 bit) field is a counter used to limit packet lifetimes. It is supposed to count time in seconds allowing a maximum lifetime of 255 sec. It must be decremented on each hop and is supposed to be decremented multiple times when queued for a long time in the router. In practice, it just counts hops. When it hits zero, the packet is dropped and a warning packet is sent back to the source host. This feature, prevents datagrams from wandering around forever, something that otherwise might happen if the routing tables become corrupted.

Protocol (8 bits) is used when IP reaches its final destination. When, the network layer has assembled a complete datagram, it needs to know what to do with it. The *Protocol* field identifies the transport protocol the network layers needs to give it to. TCP is one possibility, but so are UDP and some others. The numbering of protocols is global across the entire Internet.

The *Header checkum* verifies the header only. Such a checkum is useful for detecting errors generated by bad memory words inside a router. This algorithm is more robust than a normal add. Note, that the *Header checksum* must be recomputed at each hop because, at least one field always changes (the *time to live* field), but tricks can be used to speed up the computation.

The *Source address and Destination IP address:* These fields carry the 32 bit IP addresses of the source and destination address. One portion of the IP address indicates the network and the other portions indicate the host (or router) on the network. The IP addresses will be described in the next section.

The *Option* filed (32 bits): This field allows an IP header to be extended to be more functional. It can carry fields that control routing, timing and security.

IP Addressing

All IP addresses are 32 bits long and are used in the *Source address and Destination address* fields of IP packets.

In addition, to the physical addresses (contained on NICs) that identify individual devices, the Internet requires an additional addressing convention, an address that identifies the connection of a host to its network. Before discussing the IP addressing, let us discuss, a host and a router. A router is, fundamentally different from a host whose job is receive a datagram on an incoming link from a host and forward it on some outgoing link. Both router and a host are connected to a link through an interface. 4 Routers has multiple interfaces whereas, a host has a single interface. Because every host and a router is capable of sending and receiving IP datagram, IP requires each host and router interface to have its own IP address. Thus, an IP address is technically associated with an interface rather than with the host or a router.

For several decades, IP addresses were divided into the five categories given in the *Figure*. The different classes are designed to cover the needs of different types of organisations.

The three main address classes are class A, class B, and class C. By examining the first few bits of an address, IP software can quickly determine the class, and therefore, the structure, of an address. IP follows these rules to determine the address class:

- **Class A:** If, the first bit of an IP address is 0, it is the address of a class A network. The first bit of a class A address identifies the address class.

The next 7 bits identify the network, and the last 24 bits identify the host. There are fewer than 128 class A network numbers, but each class A network can be composed of millions of hosts.

- **Class B:** If, the first 2 bits of the address are 1 0, it is a class B network address. The first 2 bits identify class; the next 14 bits identify the network, and the last 16 bits identify the host. There are thousands of class B network numbers and each class B network can contain thousands of hosts.
- **Class C:** If, the first 3 bits of the address are 1 1 0, it is a class C network address. In a class C address, the first 3 bits are class identifiers; the next 21 bits are the network address, and the last 8 bits identify the host. There are millions of class C network numbers, but each class C network is composed of fewer than 254 hosts.
- **Class D:** If, the first 4 bits of the address are 1 1 1 0, it is a multicast address. These addresses are sometimes called class D addresses, but they don't really refer to specific networks. Multicast addresses are used to address groups of computers together at moment in time. Multicast addresses, identify a group of computers that share a common application, such as a video conference, as opposed to a group of computers that share a common network.
- **Class E:** If, the first four bits of the address are 1 1 1 1, it is a special reserved address. These addresses are called class E addresses, but they don't really refer to specific networks. No numbers are currently assigned in this range.

IP addresses are usually written as four decimal numbers separated by dots (periods). Each of the four numbers is in the range 0-255 (the decimal values possible for a single byte). Because the bits that identify class are contiguous with the network bits of the address, we can lump them together and look at the address as composed of full bytes of network address and full bytes of host address. If the value of the first byte is:

- Less than 128, the address is class A; the first byte is the network number, and the next three bytes are the host address.
- From 128 to 191, the address is class B; the first two bytes identify the network, and the last two bytes identify the host.
- From 192 to 223, the address is class C; the first three bytes are the network address, and the last byte is the host number.
- From 224 to 239, the address is multicast. There is no network part. The entire address identifies a specific multicast group.
- Greater than 239, the address is reserved.

The following table depicts each class range with other details.

Table 2: IP address classes in dotted decimal format with their ranges

IP Address Class	High Order Bit (s)	Format	Range	No. of Network Bits	No. of Host Bits	Max. Hosts	Purpose
A	0	N.H.H.H	1.0.0.0 to 126.0.0.0	7	24	2^{24}-2	Few large organisations
B	1,0	N.N.H.H	128.1.0.0 to 191.254.0.0	14	16	2^{16}-2	Medium-size organisations
C	1,1,0	N.N.N.H	192.0.1.0 to 223.255.254.0	21	8	2^{8}-2	Relatively small organisations
D	1,1,1,0	N/A	224.0.0.0 to 239.255.255.255	N/A	N/A	N/A	Multicast groups (RFC 1112)
E	1,1,1,1	N/A	240.0.0.0 to 254.255.255.255	N/A	N/A	N/A	Future Use (Experimental)

The IP address, which provides, universal addressing across all the networks of the Internet, is one of the great strengths of the TCP/IP protocol suite. However, the original class structure of the IP address has weaknesses. The TCP/IP designers did not envision the enormous scale of today's network. When TCP/IP was being designed, networking was limited to large organisations that could afford substantial computer systems. The idea of a powerful UNIX system on every desktop did not exist. At that time, a 32-bit address seemed so large that it was divided into classes to reduce the processing load on routers, even though dividing the address into classes sharply reduced the number of host addresses actually available for use. For example, assigning a large network a single class B address, instead of six class C addresses, reduced the load on the router because the router needed to keep only one route for that entire organisation. However, an organisation that was given the class B address probably did not have 64,000 computers, so most of the host addresses available to the organisation were never assigned.

The class-structured address design was critically strained by the rapid growth of the Internet. At one point it appeared that all class B addresses might be rapidly exhausted. To prevent this, a new way of looking at IP addresses without a class structure was developed.

Subnet Masks and CIDR Networks (Classless IP Addresses)

IP addresses are actually 32-bit binary numbers. Each 32-bit IP address consists of two subaddresses, one identifying the network and the other identifying the host to the network, with an imaginary boundary separating the two. The location of the boundary between the network and host portions of an IP address is determined through the use of a subnet mask. A subnet mask is another 32-bit binary number, which acts like a filter when applied to the 32-bit IP address. By comparing a subnet mask with an IP address, systems can determine the portion of the IP address that relates to the network, and the portion that relates to the host. Wherever, the subnet mask has a bit set to "1", the underlying bit in the IP address is part of the network address. Wherever the subnet mask is set to "0", the related bit in the IP address is part of the host address. For example, assume that the IP address 11000000101010000000000100010100 has a subnet mask of 11111111111111111111111100000000. In this example, the first 24 bits of the 32-bit IP addresses are used to identify the network, while the last 8 bits are used to identify the host on that network.

The size of a network (i.e., the number of host addresses available for use on it) is a function of the number of bits used to identify the host portion of the address. If, a subnet mask shows that 8 bits are used for the host portion of the address block, a maximum of 256 possible host addresses are available for that specific network. Similarly, if a subnet mask shows that 16 bits are used for the host portion of the address block, a maximum of 65,536 possible host addresses are available for use on that network.

If a network administrator needs to split a single network into multiple virtual networks, the bit-pattern in use with the subnet mask can be changed to allow as many networks as necessary. For example, assume that we want to split the 24-bit 192.168.10.0 network (which allows for 8 bits of host addressing, or a maximum of 256 host addresses) into two smaller networks. All we have to do in this situation is, change the subnet mask of the devices on the network so that they use 25 bits for the network instead of 24 bits, resulting in two distinct networks with 128 possible host addresses on each network. In this case, the first network would have a range of network addresses between 192.168.10.0 -192.168.10.127, while the second network would have a range of addresses between 192.168.10.128 -192.168.10.255.

Networks can also be enlarged through the use of a technique known as "**supernetting**," which works by extending the host portion of a subnet mask to the left, into the network portion of the address. Using this technique, a pair of networks with 24-bit subnet masks can be turned into a single large network with a 23-bit subnet mask. However, this works only if you have two neighbouring 24-bit network blocks, with the lower network having an even value (when the network portion of the address is shrunk, the trailing bit from the original network portion of the subnet mask should fall into the host portion of the new subnet mask, so the new network mask will consume both networks). For example, it is possible to combine the 24-bit 192.168.10.0 and 192.168.11.0 networks together since the loss of the trailing bit from each network (00001010 vs. 00001011) produces the same 23-bit subnet mask (0000101x), resulting in a consolidated 192.168.10.0 network. However, it is not possible to combine the 24-bit 192.168.11.0 and 192.168.12.0 networks, since the binary values in the seventh bit position (00001011 vs. 00001100) do not match when the trailing bit is removed.

Classless Inter-Domain Routing

In the modern networking environment defined by RFC 1519 [Classless Inter-Domain Routing (CIDR)], the subnet mask of a network is typically annotated in written form as a "slash prefix" that trails the network number. In the subnetting example in the previous paragraph, the original 24-bit network would be written as 192.168.10.0/24, while the two new networks would be written as 192.168.10.0/25 and 192.168.10.128/25. Likewise, when the 192.168.10.0/24 and 192.168.11.0/24 networks were joined together as a single supernet, the resulting network would be written as 192.168.10.0/23. Note, that the slash prefix annotation is generally used for human benefit; infrastructure devices still use the 32-bit binary subnet mask internally to identify networks and their routes. All networks must reserve host addresses (made up entirely of either ones or zeros), to be used by the networks themselves. This is so that, each subnet will have a network-specific address (the all-zeroes address) and a broadcast address (the all-ones address). For example, a /24 network allows for 8 bits of host addresses, but only 254 of the 256 possible addresses are available for use. Similarly, /25 networks have a maximum of 7 bits for host addresses, with 126 of the 128 possible addresses available (the all-ones and all-zeroes addresses from each subnet must be set aside for the subnets themselves). All the systems on the same subnet must use the same subnet mask in order to communicate with each other directly. If, they use different subnet masks they will think they are on different networks, and will not be able to communicate with each other without going through a router first. Hosts on different networks can use different subnet maks, although the routers will have to be aware of the subnet masks in use on each of the segments.

Subnet masks are used only by systems that need to communicate with the network directly. For example, external systems do not need to be aware of the subnet masks in use on your internal networks, since those systems will route data to your network by way of your parent network's address block. As such, remote routers need to know only the provider's subnet mask. For example, if you have a small network that uses only a /28 prefix that is, a subset of your ISP's /20 network, remote routers need to know only about your upstream provider's /20 network, while your upstream provider needs to know your subnet mask in order to get the data to your local /28 network. The rapid depletion of the class B addresses showed that three primary address classes were not enough: class A was much too large and class C was much too small. Even a class B address was too large for many networks but was used because it was better than the other alternatives.

The obvious solution to the class B address crisis was to force organisations to use multiple class C addresses. There were millions of these addresses available and they were in no immediate danger of depletion. As is often the case, the obvious solution is not as simple as it may seem. Each class C address requires its own entry within the

routing table. Assigning thousands or millions of class C addresses would cause the routing table to grow so rapidly that the routers would soon be overwhelmed. The solution requires a new way of assigning addresses and a new way of looking at addresses.

Originally network addresses were assigned in more or less sequential order as they were requested. This worked fine when the network was small and centralised. However, it did not take network topology into account. Thus, only random chance would determine if the same intermediate routers would be used to reach network 195.4.12.0 and network 195.4.13.0, which makes it difficult to reduce the size of the routing table. Addresses can only be aggregated if they are contiguous numbers and are reachable through the same route. For example, if addresses are contiguous for one service provider, a single route can be created for that aggregation because that service provider will have a limited number of routes to the Internet. But if one network address is in France and the next contiguous address is in Australia, creating a consolidated route for these addresses will not work.

Today, large, contiguous blocks of addresses are assigned to large network service providers in a manner that better reflects the topology of the network. The service providers then allocate chunks of these address blocks to the organisations to which they provide network services. This alleviates the short-term shortage of class B addresses and, because the assignment of addressees reflects the topology of the network, it permits route aggregation. Under this new scheme, we know that network 195.4.12.0 and network 195.4.13.0 are reachable through the same intermediate routers. In fact, both these addresses are in the range of the addresses assigned to Europe, 194.0.0.0 to 195.255.255.255. Assigning addresses that reflect the topology of the network enables route aggregation, but does not implement it. As long as network 195.4.12.0 and network 195.4.13.0 are interpreted as separate class C addresses, they will require separate entries in the routing table. A new, flexible way of defining addresses is therefore, needed.

Evaluating addresses according to the class rules discussed above limits the length of network numbers to 8, 16, or 24 bits - 1, 2, or 3 bytes. The IP address, however, is not really byte-oriented. It is 32 contiguous bits. A more flexible way to interpret the network and host portions of an address is with a bit mask. An address bit mask works in this way: if a bit is on in the mask, that equivalent bit in the address is interpreted as a network bit; if a bit in the mask is off, the bit belongs to the host part of the address. For example, if address 195.4.12.0 is interpreted as a class C address, the first 24 bits are the network numbers and the last 8 bits are the host addresses. The network mask that represents this is 255.255.255.0, 24 bits on and 8 bits off. The bit mask that is derived from the traditional class structure is called the default mask or the natural mask.

However, with bit masks we are no longer limited by the address class structure. A mask of 255.255.0.0 can be applied to network address 195.4.0.0. This mask includes all addresses from 195.4.0.0 to 195.4.255.255 in a single network number. In effect, it creates a network number as large as a class B network in the class C address space. Using bit masks to create networks larger than the natural mask is called supernetting, and the use of a mask instead of the address class to determine the destination network is called Classless Inter-Domain Routing (CIDR).

Specifying both the address and the mask is cumbersome when writing out addresses. A shorthand notation has been developed for writing CIDR addresses. Instead of writing network 172.16.26.32 with a mask of 255.255.255.224, we can write 172.16.26.32/27. The format of this notation is address/prefix-length, where prefix-length is the number of bits in the network portion of the address. Without this notation, the address 172.16.26.32 could easily be interpreted as a host address. RFC

1878 list all 32 possible prefix values. But little documentation is needed because the CIDR prefix is much easier to understand and remember than address classes. I know that 10.104.0.19 is a class A address, but writing it as 10.104.0.19/8 shows me that this address has 8 bits for the network number and therefore, 24 bits for the host number. I don't have to remember anything about the class A address structure.

Internet-Legal Versus Private Addressing

Although the pool of IP addresses is somewhat limited, most companies have no problems obtaining them. However, many organisations have already installed TCP/IP products on their internal networks without obtaining "legal" addresses from the proper sources. Sometimes these addresses come from example books or are simply picked at random (several firms use networks numbered 1.2.3.0, for example). Unfortunately, since they are not legal, these addresses will not be usable when these organisations attempt to connect to the Internet. These firms will eventually have to reassign Internet-legal IP addresses to all the devices on their networks, or invest in address-translation gateways that rewrite outbound IP packets so they appear to be coming from an Internet-accessible host.

Even if an address-translation gateway is installed on the network, these firms will never be able to communicate with any site that is a registered owner of the IP addresses in use on the local network. For example, if you choose to use the 36.0.0.0/8 address block on your internal network, your users will never be able to access the computers at Stanford University, the registered owner of that address block. Any attempt to connect to a host at 36.x.x.x will be interpreted by the local routers as a request for a local system, so those packets will never leave your local network.

Not all firms have the luxury of using Internet-legal addresses on their hosts, for any number of reasons. For example, there may be legacy applications that use hardcode addresses, or there may be too many systems across the organisation for a clean upgrade to be successful. If you are unable to use Internet-legal addresses, you should at least be aware that there are groups of "private" Internet addresses that can be used on internal networks by anyone. These address pools were set-aside in RFC 1918, and therefore, cannot be "assigned" to any organisation. The Internet's backbone routers are configured explicitly not to route packets with these addresses, so they are completely useless outside an organisation's internal network. The address blocks available are listed in *Table3*.

Table 3: Private Addresses Provided in RFC 1918

Class	Range of Addresses
A	Any addresses in 10.x.x.x
B	Addresses in the range of 172.16.x.x-172.31.x.x
C	Addresses in the range of 192.168.0.x-192.168.255.x

Since these addresses cannot be routed across the Internet, you must use an address-translation gateway or a proxy server in conjunction with them. Otherwise, you will not be able to communicate with any hosts on the Internet.

An important note here is that, since, nobody can use these addresses on the Internet, it is safe to assume that anybody who is using these addresses is also utilising an address-translation gateway of some sort. Therefore, while you will never see these addresses used as destinations on the Internet, if your organisation establishes a private connection to a partner organisation that is using the same block of addresses that you are using, your firms will not be able to communicate on the Internet. The packets destined for your partner's network will appear to be local to your network, and will never be forwarded to the remote network.

There are many other problems that arise from using these addresses, making their general usage difficult for normal operations. For example, many application-layer protocols embed addressing information directly into the protocol stream, and in order for these protocols to work properly, the address-translation gateway has to be aware of their mechanics. In the preceding scenario, the gateway has to rewrite the private addresses (which are stored as application data inside the application protocol), rewrite the UDP/TCP and IP checksums, and possibly rewrite TCP sequence numbers as well. This is difficult to do even with simple and open protocols such as FTP, and extremely difficult with proprietary, encrypted, or dynamic applications (these are problems for many database protocols, network games, and voice-over-IP services, in particular). These gateways almost never work for all the applications in use at a specific location.

It is always better to use formally-assigned, Internet-legal addresses whenever possible, even if, the hosts on your network do not necessarily require direct Internet access. In those cases in which your hosts are going through a firewall or application proxy of some sort, the use of Internet-legal addresses causes the least amount of maintenance trouble over time. If, for some reason this is not possible, use one of the private address pools described in *Table 3*. Do not use random, self-assigned addresses if you can possibly avoid it, as this will only cause connectivity problems for you and your users.

Fragmentation

Fragmentation is process of breaking bigger IP datagrams into smaller fragments. Each network imposes some maximum size on its packets. Maximum payloads range from 48 bytes (ATM cells) to 65,515 bytes (IP packets), although the payload size in higher layers is often bigger.

What happens if the original host sends a source packet which is too large to be handled by the destination network? The routing algorithm can hardly bypass the destination.

Basically, the only solution to the problem is to allow routers to break up packets into **fragments.**

The data in the IP datagram is broken among two or more smaller IP datagrams and these smaller fragments are then sent over the outgoing link.

The real problems in the fragmentation process is **reassembly** of fragments into a single IP datagram before sending it to the transport layer. If it is done at the network level, it will decrease the performance of a router. Therefore, to keep the network layer simple. IP V_4 designer has left it to be done at the receiver level. When the destination host receives a series of fragments from the same host, it examines the identification number of a fragment, its flag number (the flag bit is set to 0 for the last fragment whereas it is a set to 1 for all the other fragments). This is required because one or more fragments may never reach because IP provides best efforts service. **Offset** field is used to determine where the fragment fits within the original IP datagram [*Ref.5*]. When all the packets have arrived, the destination host reassembles them and sends it to the transport layer. The following table illustrates an example of 5000 bytes (20 bytes of IP header plus 4980 bytes of IP Payload arrives at a router which must be forwarded to the outgoing link with a maximum transfer unit of 1500 bytes (equivalent to Ethernet frame size).

Table 4: Fragmentation Table

Fragment	Bytes	ID	Offset	Flag
1st Fragment	1,480 bytes in the data field of the IP datagram	Identification = 999	Offset = 0 (meaning the data should be inserted beginning at byte 0)	Flag = 1 (meaning there is more)
2nd Fragment	1,480 byte information field	Identification = 999	Offset = 1,480 (meaning the data should be inserted beginning at byte 1,480)	Flag = 1 (meaning there is more)
3rd Fragment	1,480 byte information field	Identification = 999	Offset = 2,960 (meaning the data should be inserted beginning at byte 2,960)	Flag = 1 (meaning there is more)
4th Fragment	220 bytes (=4,980–1,480–1,480–1,480)	Identification = 999	Offset = 4,440 (meaning the data should be inserted beginning at byte 4, 440)	Flag = 0 (meaning this is the last fragment

This means that 4,980 data bytes in the original datagram must be allocated to four separate segments (each of which are also IP datagram). The original datagram has been stamped with an identification number 999. It is desirable to have a minimum number of fragments because fragmentation and reassembling creates extra overheads or a network system and a host. This is done by limiting the size of UDP and TCP segments to small size.

4.3.2 Internet Control Message Protocol (ICMP)

It is a protocol used by hosts and routers to send notification of datagram problems. The most typical use of ICMP is for error reporting. We have encountered errors such as, "Destination network unreachable" while running a telnet, FTP or HTTP session. This type of error reporting is done by ICMP [*Ref.5*]. As we are aware, IP is an unreliable and connectionless protocol, due to which, very often, a datagram may not reach the destination because of to a link failure, congestion etc. ICMP does reporting of such cases to the sender.

ICMP is often considered part of IP but architecturally lies just above IP as ICMP messages are carried inside IP packets. Similar to TCP and UDP segments which are carried as IP payloads, ICMP messages are also carried as IP payloads. Note, that, a datagram carries only the addresses of the original sender and the final destination. It does not know the addresses of the previous router (s) that passed a message. For this reason, ICMP can send messages only to the source and not to an intermediate router. Student are required to refer to *Reference [1] & [5]* for details of message types.

4.3.3 OSPF: The Interior Gateway Routing Protocol

Open Shortest Path First (OSPF) has become standard interior gateway routing protocol. It supports many advanced features [*Ref.1* and *Ref.5*] to meet a long list of requirements.

- The protocol specification should be publicly available. **O** in OSPF stands for **open**. It means that OSPF is not a proprietary protocol.

- It is a **dynamic algorithm** one that adapts to changes in the topology automatically and quickly.

- **Load distribution:** When multiple paths to a destination have the same cost OSPF allows multiple paths to be used.

- **Support for hierarchy** within a single routing domain. By 1988, the internet had grown so large that no router was expected to know the entire topology. OSPF was designed so that no router would have to do so.

- **Security:** All exchanges between OSPF routers are authenticated and allows only trusted routers to participate. OSPF supports three kinds of connections and networks [*Ref.1*]:

  (i) Point-to-point lines between two routers only.
  (ii) Multi-access networks with broadcasting (e.g., LANs).
  (iii) Multi-access networks without broadcasting (e.g., Packet Switched WANs).

OSPF identifies four types of routers:

(i) Internal routers (within one area).
(ii) Area border routers (connects two or more areas).
(iii) Backbone routers (performs routing within the backbone).
(iv) AS boundary routers (exchanges routing information with routers in other ASes).

OSPF allows a longer AS to be divided into smaller areas. Topology and details of one area is not visible to others. Every AS has a backbone area (called Area 0) as shown in *Figure3*.

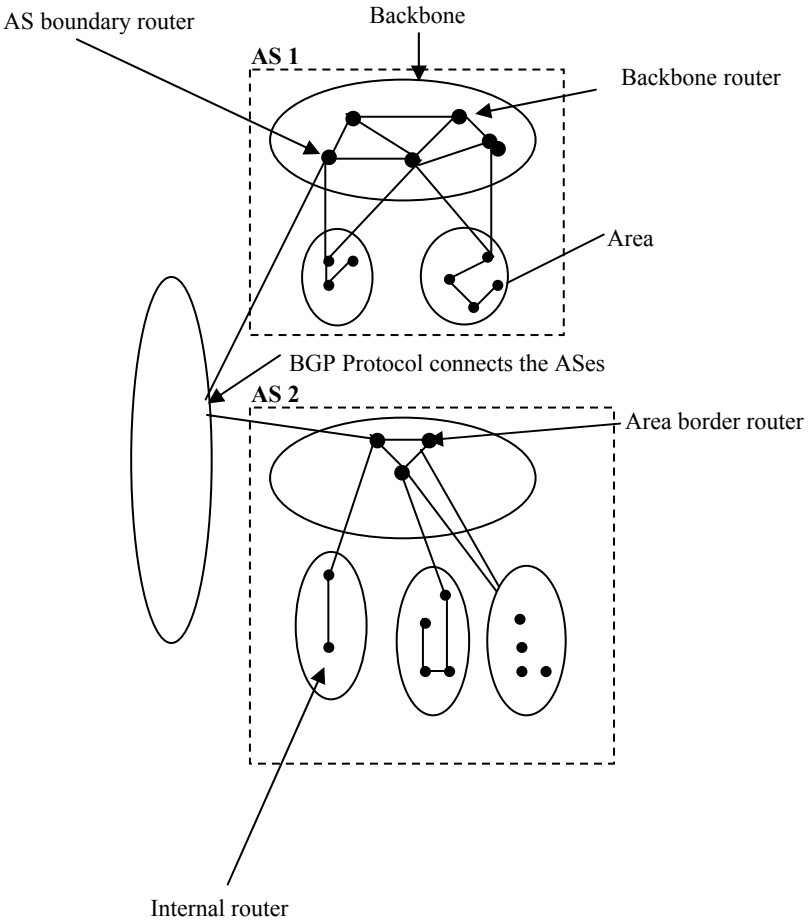

Figure 3 : Working domain of OSPF

All areas are connected to the backbone possibly by a **tunnel**, so it is possible to move from one area to another area through a backbone. The primary role of the backbone

area is to route traffic between the other areas in the AS. Inter area routing within the AS requires the follows movement of packets as shown below by arrows.

Host ⟶ internal router of an area ⟶ area border router of a source area ⟶ backbone ⟶ area border router of a destination area ⟶ internal router of a destination area ⟶ final destination

After having described all the components of OSPF, let us now conclude the topic by describing its operation.

At its heart, however, OSPF is a link state protocol that uses flooding of link state information and Dijkstra's Least-Cost path algorithm. Using flooding each router informs all other routers in its area of its neighbours and costs. This information allow each router to construct the graph for its area (s) and computers the shortest path using Dijkstra's algorithm. This is done by backbone routers also. In addition backbone routers accept information from the area border routers in order to compute the best route from each backbone router to every other router. This information is propagated back to area border routers, which advertise it within the their areas. In this manner, the optimal route is selected.

4.3.4 BGP: The Exterior Gateway Routing Protocol

The purpose of Border Gateway Protocol is to enable two different ASes to exchange routing information so that, IP traffic can flow across the AS border. A different protocol is needed between the ASes because the objectives of an interior gateway and exterior gateway routing protocol are different. Exterior gateway routing protocol such as BGP is related to policy matters. BGP is fundamentally a distance vector protocol but, it is more appropriately characterised as **path vector protocol** [*Ref.5*]. Instead of maintaining just the cost to each destination, each BGP router keeps track of the path used [*Ref.1*]. Neighbouring **BGP routers**, known as **BGP peers** exchange detailed information alongwith the list of ASes on a path to a given destination rather than record cost information.

The main advantage of using BGP is to solve the **count to infinity problem** which is illustrated in the following *Figure 4.*

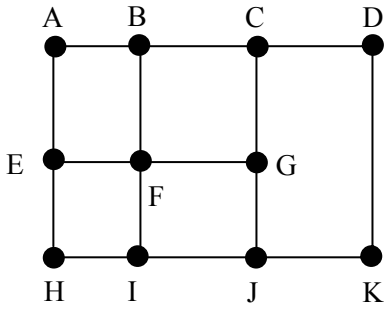

Figure 4 : Solution to Count to infinity problems in BGP

In this *Figure 4* there are A, B, C, D, E, F, G, H, I, J and K routers.
Now consider G's routing table. G uses G C D K path to forward a packet to K. As discussed earlier whenever a router gives any routing information, it provides a complete path.

For ex. From A, the path used to send a packet to K is ABCDK
From B-the path used is BCDK
From C-the path used is CGJK
From E-EFGJK
From H-HIJK.

After receiving all the paths from the neighbours, G will find the best route available. It will outright reject the path from C and E, since they pass through G itself. **Therefore, the choice left is between a route announced by B and H.** BGP easily solves count to **infinity problems**. Now, suppose C crashes or the line B-C is down. Then if B receives, two routes from its 2 neighbours: ABCDK and FBCDK, then these which can be rejected because it passes through C itself. Other distance vector algorithms make the wrong choice because, they cannot tell which of their neighbours have independent routes to their destination or not.

☞ Check Your Progress 2

1) What are the important categories of Network layer protocols?

……

……

……

2) List the important OSPF routers and their purposes.

……

……

……

3) How is BGP different from other distance vector routing protocols?

……

……

……

4.4 SUMMARY

In this unit, we defined, a large number of concepts and protocols. Tunneling is used when the source and destination networks are identical but the network, which lies in between, is different. A subnet allows a network to be split into several parts for Internet use but still acts like a single network to the outside world. It makes management of IP addresses simpler. Another reason for creating a subnet is to, establish security between different work groups. Internet is made of a large number of autonomous regions controlled by a different organisation which can, use its own routing algorithm inside. A routing algorithm within an autonomous region (such as LAN, WAN) is called an **interior gateway protocol**, an algorithm for routing between different autonomous regions are called **exterior gateway routing protocols.**

4.5 SOLUTIONS/ANSWERS

Check Your Progress 1

1) The following are the important features in which one network differs form another
   - Protocols
   - Addressing mechanism
   - Size of a packet
   - Quality of service
   - Flow control
   - Congestion control.

2) There are two such mechanisms:

   (i) Concatenated Virtual Circuit
   (ii) Datagram Approach.

3) Tunneling is used when the source and destination networks are the same but the network which lies in between is different. It uses a mechanism called encapsulation, where data transfer unit of one protocol is enclosed inside a different protocol.

Check Your Progress 2

1) There are three categories of network layer protocols:
   - IP which defines network layer addressing the field in the datagram and the action taken by routers and end systems on a datagram based on the values in these fields.
   - RIP, OSPF and BGP : They are used for routing purposes.
   - ICMP : Mainly used for error reporting in datagrams.

2) OSPF distinguishes four classes of routers:
   - Internal routers that are wholly within one area
   - Area border routers that connect two or more areas
   - Backbone routers that are on the backbone, and
   - AS boundary routers that talk to routers in other ASes.

3) It is different from other BGP protocol because its router keeps track of the path used instead of maintaining just the cost to each destination. Similarly, instead of periodically giving each neighbour its estimated cost to each possible destination, each BGP router tells its neighbours the exact path it is using.

4.6 FURTHER READINGS

1) *Computer Networks,* 4th Edition, A.S. Tanenbaum, Prentice Hall of India, New Delhi.

2) *Communication Networks fundamental concepts and key architecture,* Leon Garcia and Indra Widjaja, Tata McGraw Hill, New Delhi.

3) *Data Communication and Networking,* 2nd edition, Behrouz A. Forouzan, Tata McGraw Hill, New Delhi.

4) *Networks,* 2nd Edition, Timothy S Ramteke, Pearson Edition, New Delhi.

5) *Computer Networking A Top down Approach featuring the Internet*, James F. Kurose and Keith W. Ross Pearson Edition, New Delhi.