

Computer Networks

ECAP256

Edited by
Ajay Kumar Bansal



L OVELY
P ROFESSIONAL
U NIVERSITY



Computer Networks

**Edited By:
Ajay Kumar Bansal**

CONTENT

Unit 1:	Introduction to Computer Networks	1
	<i>Amit Sharma, Lovely Professional University</i>	
Unit 2:	Data Communication	20
	<i>Amit Sharma, Lovely Professional University</i>	
Unit 3:	<i>Network Models</i>	35
	<i>Amit Sharma, Lovely Professional University</i>	
Unit 4:	Physical Layer	56
	<i>Amit Sharma, Lovely Professional University</i>	
Unit 5:	Data Link Layer Error Detection and Correction Methods	78
	<i>Amit Sharma, Lovely Professional University</i>	
Unit 6:	Data Link layer Flow and Error Control Protocols	100
	<i>Amit Sharma, Lovely Professional University</i>	
Unit 7:	Data Link Layer Medium Access Control	120
	<i>Amit Sharma, Lovely Professional University</i>	
Unit 8:	Network layer - Logical Addressing	137
	<i>Amit Sharma, Lovely Professional University</i>	
Unit 9:	Network Layer - Routing	163
	<i>Amit Sharma, Lovely Professional University</i>	
Unit 10:	Transport layer – Protocols	191
	<i>Amit Sharma, Lovely Professional University</i>	
Unit 11:	Introduction to Computer Networks	208
	<i>Amit Sharma, Lovely Professional University</i>	
Unit 12:	Application Layer – Services and Protocols	227
	<i>Amit Sharma, Lovely Professional University</i>	
Unit 13:	Introduction to Computer Networks	249
	<i>Amit Sharma, Lovely Professional University</i>	
Unit 14:	Network Security	270
	<i>Amit Sharma, Lovely Professional University</i>	

Unit 1: Introduction to Computer Networks

CONTENTS

Objectives

Introduction

- 1.1 What is Communication?
- 1.2 Components of Data Communication System
- 1.3 Network
- 1.4 Type of Shared resources
- 1.5 Network
- 1.6 Applications of Computer Networks
- 1.7 Social Issues Originated Due to Computer Networks
- 1.8 Types of Network:
- 1.9 What is Topology?

Summary

Keywords

Self-Assessment

Objectives

After this lecture, you would be able to:

- understand the basics of network,
- learn characteristics of effective communication,
- analyze the components required for communication,
- Identify the different ways of representing the information.
- describe the main uses of computer networks
- Learn different types of network.
- Identify how one network differ from another network.
- understand how topologies can lay out a networks in different ways
- learn the various topologies as well as the various considerations in choosing the most appropriate topology

Introduction

The merging of computers and communications has a profound influence on the way systems are organized. The concept of computer center as a room with a large computer to which the users bring their work for processing is now obsolete. The old model of a single computer servicing all the computational needs of an organization has been replaced by the one in which a large system of separate but interconnected computers do the job. These systems are called computer networks. The two computers are said to be interconnected if they are able to exchange information. The connection between the computers need not be only via a copper wire or fiber optics or microwaves.

1.1 What is Communication?

Communication is a two-way process in which a sender sends a message to the receiver via a communication medium. The process is complete after the receiver sends back a feedback for an

acknowledgement to the sender that it has successfully received the information as shown in Figure 1

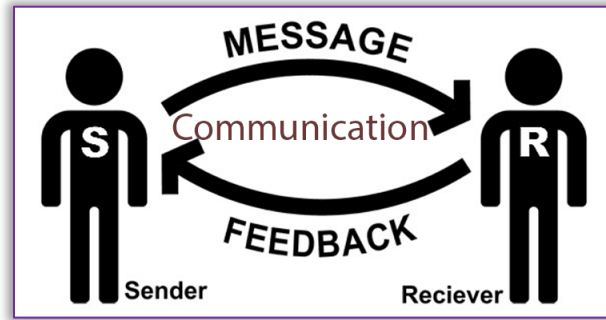


Figure 1 Process of Communication

Data Communications

Data communication is the process of transmission of digital data which could be text audio video graphics or animation between two or more computers. The communication between the two or more computers is supported by asset of wired or wireless media. This is also referred to as the transmission media. It is very important for setting up physical connection between different computers in a computer network.

Data Communication and Networking is changing the way we do business and the way we live. Business decision have to make more quickly. So, decision makers require immediate access to accurate information.

1.2 Components of Data Communication System

The various components of a data communication system could be as shown below in Figure 2 Components of Data CommunicationFigure 2.

- A. Sender: A sender is basically a computer or a device that generates a message and sends it to a receiver.
- B. Receiver: A receiver could be a computer or a device that receives a message. The receiver is generally located away from the sender and the distance between them depends upon the type of network used in between them.
- C. Message: A message could be defined as a piece of information or data that needs to be communicated from a sender to a receiver. It includes text audio video graphics or animation or a combination of any of these.
- D. Medium: A medium is a channel or a physically connected part through which a message could be transmitted from the sender to the receiver. A typical medium could be a wired or a wireless media which connects the sender and the receiver.
- E. Protocol: Protocols are simply the set of rules that govern the communication between a sender and a receiver over a network. For a communication to take place both the sender and receiver must abide by these rules.



1.3 Network

Figure 2 Components of Data Communication

A network can be defined as a group of components of a system which are closely connected together and jointly work to achieve the central objective after system. For instance, the nervous system the road system the railway lines system a few examples of interconnected components interacting together for a common objective. To understand it better latest take an example. If there is a network of agents who are selling their goods over an entire country, this would be called their marketing network. The primary objective of this group would be to enhance sales of the products and give better turnovers to the company. Ranging from the radio network the satellite network to the ancient Roman water supply network (which was used to supply water to the different parts of the Roman Empire), these networks have been an integral part of our life. This can be very well understood from the Figure 3 given below:



Figure 3 Examples of Networks

So, we can define a computer network as a group of interconnected computers and that share resources exchange files or allow communication between them.

1.4 Type of Shared resources

The different type of resources that can be shared by a computer network could be broadly categorized into two categories:

- a) **The hardware resources**
 - b) **The software resources**
- a) **The hardware resources** could be the printers, scanners, DVD ROM's or the hard disks attach to a computer which could be shared between a network of computers. It is important to note here that the sole objective of forming a network is optimal utilization of resources. The amount of resources available is always less than the total number of users using a network. So at any point of time we would not require that these resources remain underutilized or are not available to the other requesting systems. Show a computer network solve this problem by facilitating the interconnection between the different computers and resources leading to the optimal utilization of resources.
 - b) **The software resources** are the different types of files applications or software that might be used by one or more computers inside a network. These files and applications are generally the

shared resources which might be required different users for computer systems across a network.

1.5 Network

A network may be defined as a set of devices or nodes which are connected together through a wired or wireless media. These devices communicate with each other and may exchange different forms of data and instructions amongst themselves.

Computer network can be defined as an interconnection of various devices which are connected with an objective of sharing hardware software and data through wired or wireless media.

So, we can say that the computers connected together in a network can share files applications and various hardware and software resources like application programs, scanners, printers etc.

Internet can be understood as the best example of a computer network.

Basic characteristics of Computer Networks

The basic characteristics of Computer Networks are:

- a) **Fault Tolerance,**
- b) **Scalability,**
- c) **Quality of Service (QoS), and**
- d) **Security.**

Now let us understand them one by one.

- a) **Fault Tolerance:** means the ability to continue working despite failures and ensures that there is no loss of services. Suppose Raj returning home from your college with best route he knows and follows the same route every day, now one day he notices that there is a blockage in the road, so what should be done? Will you return to college or find another way to reach home? Obviously, he will find another way to reach home.

The same happens in networking, let us assume that a PC is connected with a wireless router and the router relates to router 1 and router 2 and these two routers in turn are connected with the web server. Now suppose you searched for a link or any website, but unfortunately the connection of your wireless router with router 1 is disconnected, so now wireless router will make another connection with router 2 and send the request to the web server. This is what the Fault Tolerance means.

- b) **Scalability:** means the ability to grow based on the needs and have good performance after growth. The best example of scalability is The Internet itself, now also many new users are connecting through internet and communicating with other devices, but our network is working properly.
- c) **Quality of Service (QoS):** Quality of Service (QoS) refers to the ability to set priorities and manage data traffic and reduce data loss, delay, etc.
- d) **Security:** Security is the ability to prevent unauthorized access, misuse, or forgery. Also it is the ability to provide confidentiality, integrity and availability.

Effectiveness of Data Communication

The word data refers to information presented in whatever form is agreed upon by the parties creating and using the data. Data communications are the exchange of data between two devices via some form of transmission medium such as a wire cable. For data communications to occur, the communicating devices must be part of a communication system made up of a combination of hardware (physical equipment) and software (programs). The effectiveness of a data communications system depends on four fundamental characteristics:

- a) **Delivery,**
- b) **Accuracy,**

- c) **Timeliness, and**
- d) **Jitter.**

- a) **Delivery-** The system must deliver data to the correct destination. Data must be received by the intended device or user and only by that device or user.
- b) **Accuracy-** The system must deliver the data accurately. Data that have been altered in transmission and left uncorrected are unusable.
- c) **Timeliness-** The system must deliver data in a timely manner. Data delivered late are useless. In the case of video and audio, timely delivery means delivering data as they are produced, in the same order that they are produced, and without significant delay. This kind of delivery is called real-time transmission.
- d) **Jitter-** Jitter refers to the variation in the packet arrival time. It is the uneven delay in the delivery of audio or video packets. For example, let us assume that video packets are sent every 3D ms. If some of the packets arrive with 3D-ms delay and others with 4D-ms delay, an uneven quality in the video is the result.

Uses of Data communication and computer networks

Computer networks have become invaluable to organizations as well as individuals. Some of its main uses are as follows –

- a) **Information and Resource Sharing** – Computer networks allow organizations having units which are placed apart from each other, to share information in a very effective manner. Resource sharing such a sharing of a printer, and various storage devices connected across multiple computers on the network. Programs and software in any computer can be accessed by other computers linked to the network. It also allows sharing of hardware equipment, like printers and scanners among varied users.
- b) **File sharing** - It provides access to digital media, such as a computer, programs, electronic book or documents. Common methods of storage and transmission include manual sharing utilizing removable media, centralized servers on computer networks.
- c) **Application Sharing** - It enables two or more users to access a shared application or software over the network with the help of the client/server application.
- d) **Hardware sharing** - In hardware, sharing users can access hardware devices like Printer, Hard disk, Ram, etc. with the help centralized computer or device.
- e) **User Communication** - Networks allow users to communicate using e-mail, newsgroups, and video conferencing, etc.
- f) **Network Gaming** - Network Gaming in Benefits of computer network. A user can get the Advantage of gaming over the network as it also provides network gaming where two or more users can play a game from a different location.
- g) **Retrieving Remote Information** – through computer networks, users can retrieve remote information on a variety of topics. The information is stored in remote databases to which the user gains access through information systems like the World Wide Web.
- h) **Speedy Interpersonal Communication** – computer networks have increased the speed and volume of communication like never before. Electronic Mail (email) is extensively used for sending texts, documents, images, and videos across the globe. Online communications have increased by manifold times through social networking services.
- i) **E-Commerce** – computer networks have paved way for a variety of business and commercial transactions online, popularly called e-commerce. Users and organizations can pool funds, buy,

or sell items, pay bills, manage bank accounts, pay taxes, transfer funds and handle investments electronically.

- j) **Highly Reliable Systems** – computer networks allow systems to be distributed in nature, by the virtue of which data is stored in multiple sources. This makes the system highly reliable. If a failure occurs in one source, then the system will continue to function, and data will still be available from the other sources.
- k) **Cost-Effective Systems** – computer networks have reduced the cost of establishment of computer systems in organizations. Previously, it was imperative for organizations to set up expensive mainframes for computation and storage. With the advent of networks, it is sufficient to set up interconnected personal computers (PCs) for the same purpose.
- l) **VoIP** – VoIP or Voice over Internet protocol has revolutionized telecommunication systems. Through this, telephone calls are made digitally using Internet Protocols instead of the regular analog phone lines.

Components

Computer network components are the major parts which are needed to install the software. Some important network components are NIC, switch, cable, hub, router, and modem. Depending on the type of network that we need to install, some network components can also be removed. For example, the wireless network does not require a cable.

Following are the major components required to install a network:

- a) **NIC:** NIC stands for network interface card. NIC is a hardware component used to connect a computer with another computer onto a network. It can support a transfer rate of 10,100 to 1000 Mb/s. The MAC address or physical address is encoded on the network card chip which is assigned by the IEEE to identify a network card uniquely. The MAC address is stored in the PROM (Programmable read-only memory).

There are two types of NIC as can be seen in the Figure 4:

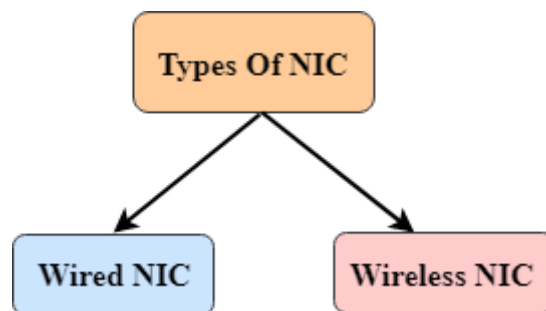


Figure 4 Types of NIC

- i **Wired NIC:** The Wired NIC is present inside the motherboard. Cables and connectors are used with wired NIC to transfer data.
 - ii **Wireless NIC:** The wireless NIC contains the antenna to obtain the connection over the wireless network. For example, laptop computer contains the wireless NIC.
- b) **Hub:** A Hub is a hardware device that divides the network connection among multiple devices. When computer requests for some information from a network, it first sends the request to the Hub through cable. Hub will broadcast this request to the entire network. All the devices will check whether the request belongs to them or not. If not, the request will be dropped.

The process used by the Hub consumes more bandwidth and limits the amount of communication. Nowadays, the use of hub is obsolete, and it is replaced by more advanced computer network components such as Switches, Routers.

- c) **Switch:** A switch is a hardware device that connects multiple devices on a computer network. A Switch contains more advanced features than Hub. The Switch contains the updated table that decides where the data is transmitted or not. Switch delivers the message to the correct destination based on the physical address present in the incoming message. A Switch does not broadcast the message to the entire network like the Hub. It determines the device to whom the message is to be transmitted. Therefore, we can say that switch provides a direct connection between the source and destination. It increases the speed of the network.
- d) **Router:** A router is a hardware device which is used to connect a LAN with an internet connection. It is used to receive, analyze and forward the incoming packets to another network. A router works in a Layer 3 (Network layer) of the OSI Reference model. A router forwards the packet based on the information available in the routing table. It determines the best path from the available paths for the transmission of the packet.

Advantages of Router:

1. **Security:** The information which is transmitted to the network will traverse the entire cable, but the only specified device which has been addressed can read the data.
 2. **Reliability:** If the server has stopped functioning, the network goes down, but no other networks are affected that are served by the router.
 3. **Performance:** Router enhances the overall performance of the network. Suppose there are 24 workstations in a network generates a same amount of traffic. This increases the traffic load on the network. Router splits the single network into two networks of 12 workstations each, reduces the traffic load by half.
 4. **Network range:** Router extends the overall range of the network.
- e) **Modem:** A modem is a hardware device that allows the computer to connect to the internet over the existing telephone line. A modem is not integrated with the motherboard rather than it is installed on the PCI slot found on the motherboard. It stands for Modulator/Demodulator. It converts the digital data into an analog signal over the telephone lines.

Based on the differences in speed and transmission rate, a modem can be classified in the following categories:

- Standard PC modem or Dial-up modem
- Cellular Modem
- Cable modem

1.6 Applications of Computer Networks

Had it not been of high importance, nobody would have bothered connecting computers over a network. Let's start exploring the uses of Computer Networks with some traditional use cases at companies and for individuals and then move on to the recent developments in the field of mobile users and home networking.

Computer Networks for Business Applications

Following are some business applications of computer networks:

- a) **Resource Sharing:** The goal is to make all programs, equipment (like printers etc), and especially data, available to anyone on the network without regard to the physical location of the resource and the user.
- b) **Server-Client model:** One can imagine a company's information system as consisting of one or more databases and some employees who need to access it remotely. In this model, the data is stored on powerful computers called Servers. Often these are centrally housed and maintained by a system administrator. In contrast, the employees have simple machines, called Clients, on their desks, using which they access remote data.
- c) **Communication Medium:** A computer network can provide a powerful communication medium among employees. Virtually every company that has two or more computers now has e-mail (electronic mail), which employees generally use for a great deal of daily communication
- d) **E-commerce:** A goal that is starting to become more important in businesses is doing business with consumers over the Internet. Airlines, bookstores, and music vendors have discovered that many customers like the convenience of shopping from home. This sector is expected to grow quickly in the future.

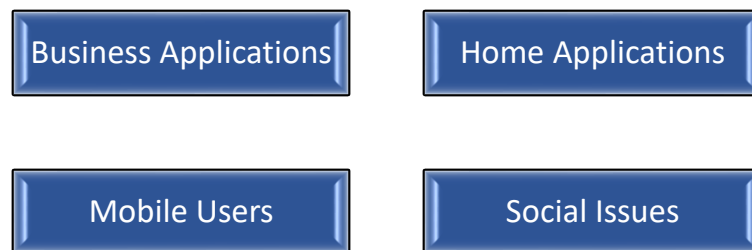


Figure 5 Applications of Computer Networks

The most popular forms are listed in the Table 1 below:

Table 1 The most popular forms of E-commerce

B2C (Business to Consumer)	When consumer order shoes online
B2B (Business to Business)	Truck manufacturer ordering tires from suppliers
C2C (Consumer to Consumer)	An online auction site(eBay)
G2C (Government to Citizen)	Reduce the average time for fulfilling citizen's requests for various government services.
P2P (Peer to Peer)	Buyer and seller transact directly

1. Computer Networks for Home Applications

Some of the most important uses of the Internet for home users are as follows:

- Access to remote information
- Person-to-person communication
- Interactive entertainment
- Electronic commerce

2. Computer Networks for Mobile Users

Mobile computers, such as notebook computers and Mobile phones, is one of the fastest-growing segments of the entire computer industry. Although wireless networking and mobile computing are often related, they are not identical.

1.7 Social Issues Originated Due to Computer Networks

The growth in the availability of affordable computing technology has caused several major shifts in the way that society operates. The majority of these have been for the better, with home computers and the internet providing unlimited access to all the information ever created and discovered by humanity.

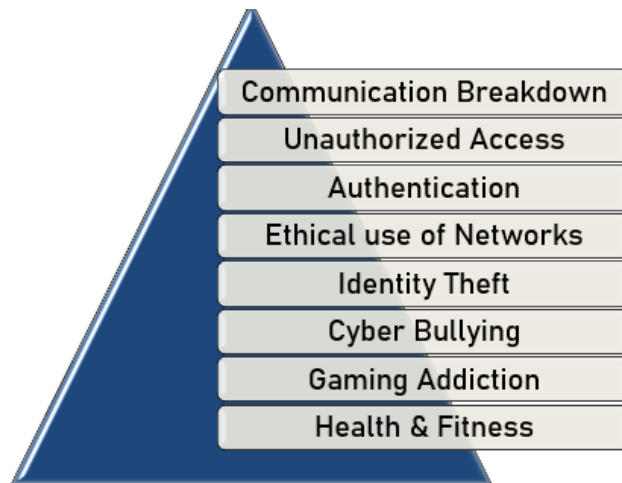


Figure 6 Social Issues of Computer Networks

There are, however, some less positive social issues generated as a direct result of technological advances as can be seen in Figure 6. In the interests of balance, it is important to analyze these and assess the severity of their impact so that steps can be taken to better understand and combat the negative effects.

1. Communication Breakdown

Socializing within a family unit has always been important, as it strengthens the bonds between us and ensures cohesion within the group. But with more and more households owning several computers and numerous portable devices granting access to information and entertainment, some argue that this is leading to a lack of family communication. If each member is engrossed in their laptop, smartphone or tablet each evening, even communal things like watching television are compromised. Meanwhile, you can see whole families who are out to dinner and still staring into a touchscreen rather than talking to one another. And if you're the one driving to that family dinner and texting while driving, you're a distracted driver, increasing your risk of crashing, and potentially causing death and injury. Increase your digital wellbeing by allowing technology to improve your life and not to become a distraction to your life and others. Your life and others are more important than technology.

2. Defamation of Character

The only means of getting in touch with major corporations or famous people in the public eye prior to the advent of digital communication was via a stiffly written letter. This was, of course, accessible only to the intended recipient and thus a very private way for the disgruntled to vent their spleen. But first message boards and now social media services like Facebook and Twitter are being used to defame people and businesses in an intrinsically public manner. This has led to arrests, lawsuits and the threat of placing stricter controls over what can and cannot be posted to such services. It has also caused heartache and woe for many individuals, helping to perpetuate a massive, international rumor mill which pays little heed to facts or the threat of legal action.

3. Identity Theft

Fraud is another spurious activity that has been able to evolve in the wake of easily accessible computers and the internet. Perhaps most problematic and prevalent of the various fraudulent activities is identity theft, in which personal details of innocent people are harvested by a third party so that they can be used for malicious purposes. This includes carrying out illicit online transactions and other damaging activities that can have serious ramifications.

4. Cyber Bullying

As with the defamation of public figures, the internet and computers have also made it easier for spiteful people to attack people they know personally as well as perfect strangers via the anonymous platforms that are available to them. This has led to serious incidents of cyber bullying involving both children and adults, sometimes with tragic consequences. The problem with these techniques is that they tend to go under the radar to an even greater degree than traditional bullying, which makes it harder to detect and correct.

5. Gaming Addiction

Whilst computers and the internet have made it easier for gambling addicts to get their fix, a new type of addiction has also arisen, in the form of addiction to videogames. This is something that can impact people of all ages and leads inevitably to a number of problems, from the social to the financial. Professionals are beginning to take gaming addiction seriously and combat it in the same way as other diseases.

6. Privacy

Whilst high profile cases of online identity theft and fraud should have caused people to become more careful about how they use their personal information, issues of privacy and a lack of appreciation for the risks are still widespread. This extends beyond simply giving away private data via chat rooms, message boards and e-commerce sites and extends into the compromising world of social media. Employers are now combing Facebook and Twitter to effectively do background checks on potential employees, paying particular attention to those that have not chosen to use privacy settings to prevent anyone from getting a look at their details.

7. Health & Fitness

We are living increasingly sedentary lifestyles, because computers are removing the need for us to physically carry out many tasks, as well as keeping us rooted to one spot throughout our working days and during our leisure time. Not to mention the physical issues that can arise from constantly looking down at our smart devices, forever buried in a blur of FOMO (Fear of Missing Out) delirium. This is leading to an epidemic of childhood and adult obesity throughout the developed world, with the UK possessing one of the worst records in this respect of any of its Western neighbors.

8. Education

The educational properties of computers are well known and universally lauded but having all the information in existence on tap has its own issues. In particular, the practice of plagiarism has become a major problem, as students can simply copy and paste whole chunks of text from online sources without attributing the work to anyone else. This has become the bane of educational institutions, which tend to come down hard on detected plagiarists in order to discourage similar activities from others.

9. Terrorism & Crime

Computers have been a positive force in allowing for the creation of global movements and righteous activism in several forms. However, the other side of the coin is that terrorists and organized criminals also exploit the web for their own nefarious purposes. Businesses, governments, and individuals are all at risk of cyber-attack and the perpetrators can often act anonymously from a country with no extradition agreements.

1.8 Types of Network:

A computer network is a group of computers linked to each other that enables the computer to communicate with another computer and share their resources, data, and applications.

A computer network can be categorized by their size. A computer network is mainly of four types:

1) LAN (Local Area Network)

- 2) PAN (Personal Area Network)
- 3) MAN (Metropolitan Area Network)
- 4) WAN (Wide Area Network)

LAN (Local Area Network)

Local Area Network is a group of computers connected to each other in a small area such as building, office. It is used for connecting two or more personal computers through a communication medium such as twisted pair, coaxial cable, etc. It can very well be understood with the help of the Figure 7 shown below. It is less costly as it is built with inexpensive hardware such as hubs, network adapters, and ethernet cables. The data is transferred at an extremely faster rate in Local Area Network. Local Area Network provides higher security.

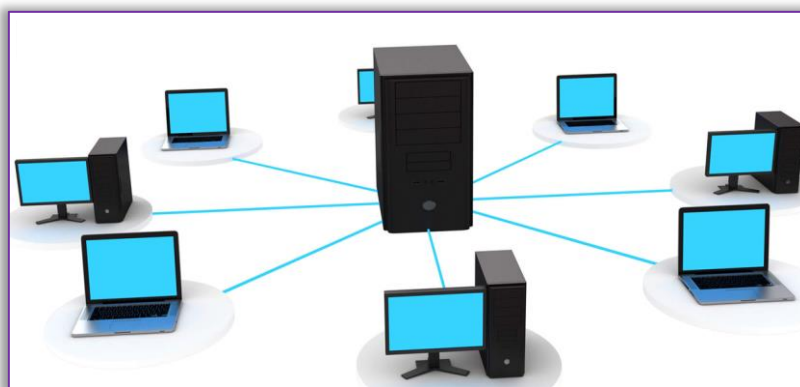


Figure 7 Types of LAN

LAN can further be of Two types:

- 1) *Client-Server Network:*
- 2) *Peer-to-Peer Network:*

Table 2 Client-Server vs Peer-to-Peer Network

S.No.	Client-Server Network	Peer-to-Peer Network
1.	In Client-Server Network, Clients and server are differentiated, Specific server and clients are present.	In Peer-to-Peer Network, Clients and server are not differentiated.
2.	Client-Server Network focuses on information sharing.	While Peer-to-Peer Network focuses on connectivity.
3.	In Client-Server Network, Centralized server is used to store the data.	While in Peer-to-Peer Network, Each peer has its own data.
4.	In Client-Server Network, Server respond the services which is request by Client.	While in Peer-to-Peer Network, Each and every node can do both request and respond for the services.
5.	Client-Server Network are costlier than Peer-to-Peer Network.	While Peer-to-Peer Network are less costlier than Client-Server Network.
6.	Client-Server Network are more stable than Peer-to-Peer Network.	While Peer-to-Peer Network are less stable if number of peer is increase.
7.	Client-Server Network is used for both small and large networks.	While Peer-to-Peer Network is generally suited for small networks with fewer than 10 computers.



6808 • J. Neurosci., September 24, 2008 • 28(39):6803–6812

Personal Area Network is a network arranged within an individual person, typically within a range of 10 meters. It is used for connecting the computer devices of personal use is known as Personal Area Network. Thomas Zimmerman was the first research scientist to bring the idea of the Personal Area Network. It covers an area of 30 feet.

Personal Area Network is a network arranged within an individual person, typically within a range of 10 meters. It is used for connecting the computer devices of personal use is known as Personal Area Network. Thomas Zimmerman was the first research scientist to bring the idea of the Personal Area Network. It covers an area of 30 feet.



There are two types of Personal Area Network:

- 1) *Wired Personal Area Network*
- 2) *Wireless Personal Area Network*

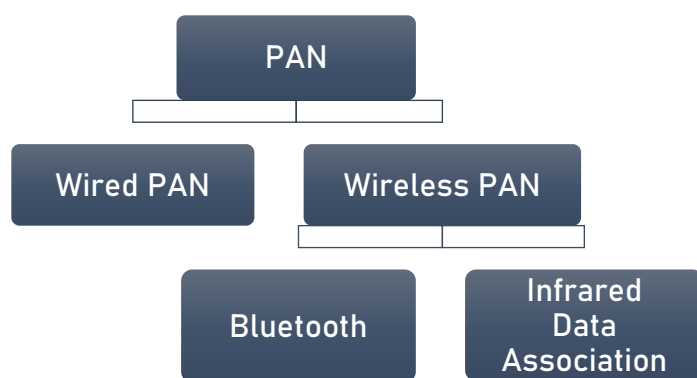


Figure 10 Types of PAN

- a) **Wireless Personal Area Network:** Wireless Personal Area Network is developed by simply using wireless technologies such as Wi-Fi, Bluetooth. It is a low range network.
- b) **Wired Personal Area Network:** Wired Personal Area Network is created by using the USB.

Examples of Personal Area Network:

- **Body Area Network:** Body Area Network is a network that moves with a person. For example, a mobile network moves with a person. Suppose a person establishes a network connection and then creates a connection with another device to share the information.
- **Offline Network:** An offline network can be created inside the home, so it is also known as a home network. A home network is designed to integrate the devices such as printers, computer, television but they are not connected to the internet.
- **Small Home Office:** It is used to connect a variety of devices to the internet and to a corporate network using a VPN

MAN (Metropolitan Area Network)

A metropolitan area network is a network that covers a larger geographic area by Interconnecting a different LAN to form a larger network. Government agencies use MAN to connect to the citizens and private industries. In MAN, various LANs are connected to each other through a telephone exchange line. The most widely used protocols in MAN are RS-232, Frame Relay, ATM, ISDN, OC-3, ADSL, etc. It has a higher range than Local Area Network (LAN).

Uses of Metropolitan Area Network:

- MAN is used in communication between the banks in a city.
- It can be used in an Airline Reservation.
- It can be used in a college within a city.
- It can also be used for communication in the military.

WAN (Wide Area Network)

- A Wide Area Network is a network that extends over a large geographical area such as states or countries.
- A Wide Area Network is quite bigger network than the LAN.
- A Wide Area Network is not limited to a single location, but it spans over a large geographical area through a telephone line, fiber optic cable or satellite links.
- The internet is one of the biggest WAN in the world.
- A Wide Area Network is widely used in the field of Business, government, and education.

Examples of Wide Area Network:

- Mobile Broadband: A 4G network is widely used across a region or country.
- Last mile: A telecom company is used to provide the internet services to the customers in hundreds of cities by connecting their home with fiber.
- Private network: A bank provides a private network that connects the 44 offices. This network is made by using the telephone leased line provided by the telecom company.

Advantages of Wide Area Network:

- Geographical area: A Wide Area Network provides a large geographical area. Suppose if the branch of our office is in a different city then we can connect with them through WAN. The internet provides a leased line through which we can connect with another branch.
- Centralized data: In case of WAN network, data is centralized. Therefore, we do not need to buy the emails, files or back up servers.
- Get updated files: Software companies work on the live server. Therefore, the programmers get the updated files within seconds.
- Exchange messages: In a WAN network, messages are transmitted fast. The web application like Facebook, WhatsApp, Skype allows you to communicate with friends.
- Sharing of software and resources: In WAN network, we can share the software and other resources like a hard drive, RAM.
- Global business: We can do the business over the internet globally.
- High bandwidth: If we use the leased lines for our company then this gives the high bandwidth. The high bandwidth increases the data transfer rate which in turn increases the productivity of our company.

Disadvantages of Wide Area Network:

- Security issue: A WAN network has more security issues as compared to LAN and MAN network as all the technologies are combined together that creates the security problem.
- Needs Firewall & antivirus software: The data is transferred on the internet which can be changed or hacked by the hackers, so the firewall needs to be used. Some people can inject the virus in our system, so antivirus is needed to protect from such a virus.
- High Setup cost: An installation cost of the WAN network is high as it involves the purchasing of routers, switches.
- Troubleshooting problems: It covers a large area so fixing the problem is difficult.

Internetwork

- An internetwork is defined as two or more computer network LANs or WAN or computer network segments are connected using devices, and they are configured by a local addressing scheme. This process is known as internetworking.
- An interconnection between public, private, commercial, industrial, or government computer networks can also be defined as internetworking.
- An internetworking uses the internet protocol.
- The reference model used for internetworking is Open System Interconnection (OSI).

Types of Internetwork:

1. **Extranet:** An extranet is a communication network based on the internet protocol such as Transmission Control protocol and internet protocol. It is used for information sharing. The access to the extranet is restricted to only those users who have login credentials. An extranet is the lowest level of internetworking. It can be categorized as MAN, WAN or other

computer networks. An extranet cannot have a single LAN, atleast it must have one connection to the external network.

2. **Intranet:** An intranet is a private network based on the internet protocol such as Transmission Control protocol and internet protocol. An intranet belongs to an organization which is only accessible by the organization's employee or members. The main aim of the intranet is to share the information and resources among the organization employees. An intranet provides the facility to work in groups and for teleconferences.

Advantages of Intranet:

- **Communication:** It provides a cheap and easy communication. An employee of the organization can communicate with another employee through email, chat.
- **Timesaving:** Information on the intranet is shared in real time, so it is timesaving.
- **Collaboration:** Collaboration is one of the most important advantage of the intranet. The information is distributed among the employees of the organization and can only be accessed by the authorized user.
- **Platform independency:** It is a neutral architecture as the computer can be connected to another device with different architecture.
- **Cost effective:** People can see the data and documents by using the browser and distributes the duplicate copies over the intranet. This leads to a reduction in the cost.

The differences between LAN MAN and WAN can be clearly understood with the help of Table3

Table 3 Comparison of LAN, WAN and MAN

LAN	MAN	WAN
LAN's ownership is private.	MAN's ownership can be private or public.	While WAN also might not be owned by one organization.
The transmission speed of a LAN is high.	The transmission speed of a MAN is average.	The transmission speed of a WAN is low.
The propagation delay is short in a LAN.	There is a moderate propagation delay in a MAN.	There is a long propagation delay in a WAN.
Less congestion in LAN	More congestion in MAN	Even more congestion than MAN in WAN
LAN's design and maintenance are easy	While MAN's design and maintenance is difficult than LAN	Whereas WAN's design and maintenance are also difficult than LAN as well MAN
There is more fault tolerance in LAN	While there is less fault tolerance	In WAN, there is also less fault tolerance

1.9 What is Topology?

Topology defines the structure of the network of how all the components are interconnected to each other. There are two types of topology: physical and logical topology.

Physical topology is the geometric representation of all the nodes in a network.

- **Physical topology:** The placement of the various components of a network, including device location and cable installation.

- **Logical topology:** illustrates how data flows within a network, regardless of its physical design.

The different types of topologies are:

1) Point-to-Point

Point-to-point networks contains exactly two hosts such as computer, switches, routers, or servers connected back to back using a single piece of cable. Often, the receiving end of one host is connected to sending end of the other and vice versa.

If the hosts are connected point-to-point logically, then may have multiple intermediate devices. But the end hosts are unaware of underlying network and see each other as if they are connected directly.

2) Bus Topology

In case of Bus topology, all devices share single communication line or cable. Bus topology may have problem while multiple hosts sending data at the same time. Therefore, Bus topology either uses CSMA/CD technology or recognizes one host as Bus Master to solve the issue. It is one of the simple forms of networking where a failure of a device does not affect the other devices. But failure of the shared communication line can make all other devices stop functioning.

Both ends of the shared channel have line terminator. The data is sent in only one direction and as soon as it reaches the extreme end, the terminator removes the data from the line.

3) Star Topology

All hosts in Star topology are connected to a central device, known as hub device, using a point-to-point connection. That is, there exists a point to point connection between hosts and hub. The hub device can be any of the following:

- Layer-1 device such as hub or repeater
- Layer-2 device such as switch or bridge
- Layer-3 device such as router or gateway

As in Bus topology, hub acts as single point of failure. If hub fails, connectivity of all hosts to all other hosts fails. Every communication between hosts takes place through only the hub. Star topology is not expensive as to connect one more host, only one cable is required and configuration is simple.

4) Ring Topology

In ring topology, each host machine connects to exactly two other machines, creating a circular network structure. When one host tries to communicate or send message to a host which is not adjacent to it, the data travels through all intermediate hosts. To connect one more host in the existing structure, the administrator may need only one more extra cable. Failure of any host results in failure of the whole ring. Thus, every connection in the ring is a point of failure. There are methods which employ one more backup ring.

5) Mesh Topology

In this type of topology, a host is connected to one or multiple hosts. This topology has hosts in point-to-point connection with every other host or may also have hosts which are in point-to-point connection with few hosts only. Hosts in Mesh topology also work as relay for other hosts which do not have direct point-to-point links. Mesh technology comes into two types:

- Full Mesh: All hosts have a point-to-point connection to every other host in the network. Thus, for every new host $n(n-1)/2$ connections are required. It provides the most reliable network structure among all network topologies.
- Partially Mesh: Not all hosts have point-to-point connection to every other host. Hosts connect to each other in some arbitrarily fashion. This topology exists where we need to provide reliability to some hosts out of all.

6) Tree Topology

Also known as Hierarchical Topology, this is the most common form of network topology in use presently. This topology imitates as extended Star topology and inherits properties of Bus topology. This topology divides the network into multiple levels/layers of network. Mainly in LANs, a network is bifurcated into three types of network devices. The lowermost is access-layer where computers are attached. The middle layer is known as distribution layer, which works as mediator between upper layer and lower layer. The highest layer is known as core layer, and is central point of the network, i.e. root of the tree from which all nodes fork. All neighboring hosts have point-to-point connection between them. Like the Bus topology, if the root goes down, then the entire network suffers even though it is not the single point of failure. Every connection serves as point of failure, failing of which divides the network into unreachable segment.

7) Daisy Chain

This topology connects all the hosts in a linear fashion. Similar to Ring topology, all hosts are connected to two hosts only, except the end hosts. Means, if the end hosts in daisy chain are connected then it represents Ring topology.

Each link in daisy chain topology represents single point of failure. Every link failure splits the network into two segments. Every intermediate host works as relay for its immediate hosts.

8) Hybrid Topology

A network structure whose design contains more than one topology is said to be hybrid topology. Hybrid topology inherits merits and demerits of all the incorporating topologies.

The above picture represents an arbitrarily hybrid topology. The combining topologies may contain attributes of Star, Ring, Bus, and Daisy-chain topologies. Most WANs are connected by means of Dual-Ring topology and networks connected to them are mostly Star topology networks. Internet is the best example of largest Hybrid topology.

Summary

- A network consists of two or more computers that are linked in order to share resources (such as printers and CD-ROMs), exchange files or allow electronic communications. The computers on a network may be linked through cables, telephone lines, radio waves, satellites or infrared light beams.
- The primary purpose of a computer network is to share resources. The main goal of networking is Resource sharing. A second goal is to provide high reliability by having alternative sources of supply. Another goal is saving money. Another closely related goal is to increase the systems performance as the work load increases by just adding more processors. With central mainframes, when the system is full, it must be replaced by a larger one, usually at great expense and with even greater disruption to the users. Computer networks provide a powerful communication medium.
- There are two important dimensions for classifying networks — transmission technology and scale.
- Transmission technology can be classified into two types:
 1. Broadcast networks.
 2. Point-to-point networks.
- Broadcast networks: These networks have a single communication channel shared by all the machines on the network.
- Point-to-point networks consist of many connections between individual pairs of machines.
- Multiple routes and intermediate machines may exist between a pair of machines; so routing algorithms play an important role here.
- A collection of interconnected networks is called an internetwork or just Internet. The Internet refers to a specific worldwide Internet that is widely used to connect universities, government offices, companies and private individuals.
- A network topology is the basic design of a computer network. It details how key network components such as nodes and links are interconnected.
- There are three primary types of network topologies which refer to the physical and logical layout of the Network cabling. They are star, ring and bus topology.

Keywords

Archive: A computer site advertises and stores a large amount of public domain, shareware software and documentation.

Broadcast Networks: They have a single communication channel, which is shared by all the computers on the network and therefore, any message transmitted by a computer on the network is received by all the computers connected to the channel.

Error Control: The receiving end after completion of receiving the information must also be capable of dealing with and recognizing the corruption.

Local Area Network: A LAN is a form of local (limited distance), shared packet network for computer communications.

Metropolitan Area Network: In MAN, different LANs are connected through a local telephone exchange using one or two cables but not switching elements.

Service Primitives: The primitives enable the service provider to perform some action or report on an action taken by a peer entity.

Wide Area Network: A WAN may be defined as a data communications network that covers a relatively broad geographic area to connect LANs together between different cities with the help of transmission facilities provided by common carriers, such as telephone companies.

Self-Assessment

State whether the following statements are true or false:

Fill in the blanks:

1. The main goal of networking is
2. In a distributed system, the existence of multiple autonomous computers is..... to the user.
3. The computers on a may be linked through cables, telephone lines, radio waves, satellites or infrared light beams.
4. 4. You can create files and store them in one computer, then those files from the other computer(s) connected to it.
5. A system is a special case of a network; one whose software gives it a high degree of cohesiveness and transparency.

State whether the following is true or false.

6. A topology, which is a pattern of interconnections among nodes, influences a network's cost and performance.
7. There are Five primary types of network topologies which refer to the physical and logical layout of the Network cabling.
8. Bus is the simplest and the oldest and all the telephone switches are based on this.
9. Bus consists of a single cable, called a Backbone that connects all workstations on the network using a single line.
10. The purpose of the terminators at either end of the network is to stop the signal being reflected back.

1.10 Review Questions

1. What are the major factors that have made the use of computer networks as an integral part of the business?
2. How are computer networks classified? Mention the some of the important reasons for the classification of computer networks.
3. How is LAN characterized? Explain.

4. What are the different technologies available for implementing WAN?
5. What is WAN? How does it differ from LANs and MANs? Give at least two examples of popular WANs.

Answers: Self Assessment

- | | |
|---------------------|----------------|
| 1. Resource sharing | 2. transparent |
| 3. network | 4. access |
| 5. distributed | 6. True |
| 7. False | 8. False |
| 9. True | 10. True |

Further Readings



Andrew S. Tanenbaum, *Computer Networks*, Prentice Hall

Behrouz A. Forouzan and Sophia Chung Fegan, *Data Communications and Networking*, McGraw-Hill Companies

Burton, Bill, *Remote Access for Cisco Networks*, McGraw-Hill Osborne Media McGraw-Hill Osborne Media

Rajneesh Agrawal and Bharat Bhushan Tiwari, *Computer Networks and Internet*, Vikas Publication

Kurose and Ross, *Computer Networking: A Top Down Approach*, Addison-Wesley, (2012), 6th ed.

Stallings, W., *Computer Networking with Internet Protocols and Tech*, Prentice Hall of India (2010), 9th ed.

Unit 02: Data Communication

CONTENT

Objectives

Introduction

2.1 What is Data Communication

2.2 Signal Classifications

2.3 Transmission Mode

2.4 Transmission Impairments

2.5 Protocols

2.6 Network Standards

Summary

Self-Assessment

Answers for Self Assessment

Review Questions

Further Readings

Objectives

After this lecture, you would be able to:

- understand analog and digital signals
- learn what are the various transmission modes.
- understand what is performance metrics.
- learn the various essential network metrics to monitor.
- Understand the various reasons for transmission impairments
- Understand the concept of protocols.
- Learn the components and functions of protocols

Introduction

Data Communication facilitates transmission of signals via. different transmission media. During this process it deals with different types of signals. Re signals can be classified based upon different classification criteria's like their ability to exhibit a pattern, or the type of signal that is transmitted from the source to the destination. This weather in walls different transmission modes and consideration of a few essential network performance metrics which decide the quality e of data transmitted within the channel. However, there are a few challenges posed by the different transmission impairments.

2.1 What is Data Communication

Data communication is the process of transmission of digital data which could be text audio video graphics or animation between two or more computers. The communication between the two or more computers is supported by asset of wired or wireless media. This is also referred to as the transmission media. It is very important for setting up physical connection between different computers in a computer network.

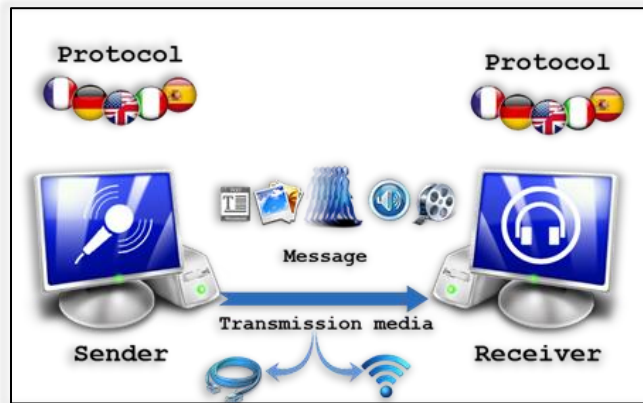


Figure 1 Data Communication Process

2.2 Signal Classifications

Signals can be classified based on different parameters like the time intervals between which it is being created, or the type of the signal.

a) Periodic and Non-Periodic Signals

Periodic Signals	Non-Periodic Signals
Completes a pattern and repeats that pattern over subsequent identical periods.	Changes without exhibiting a pattern
The completion of one full pattern is called a cycle.	Any continuous-time signal which is not periodic is called a non-periodic signal.

b) Analog and Digital Data

i Analog Signal

- Many levels of intensity
- An infinite number of values.

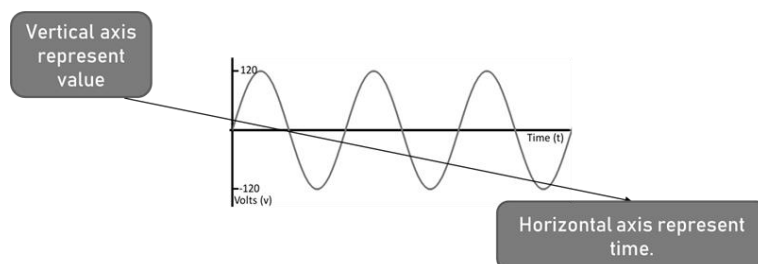


Figure 2 Analog Signal

ii Digital Signal

- Limited number of defined values.
- Although each value can be any number, it is often as simple as 1 and 0.

Analog Data	Digital Data
The term analog data refers to information that is continuous.	Digital data take on discrete values.
Example 	Example 

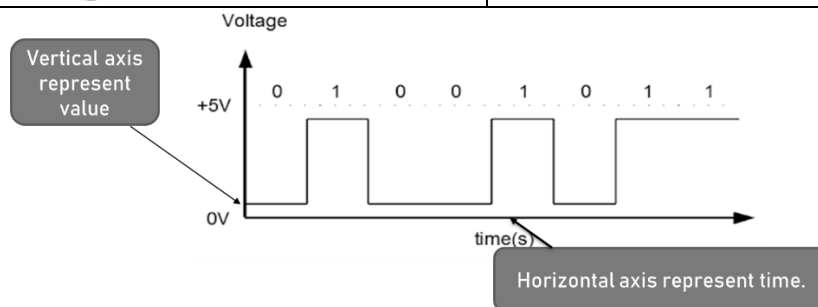


Figure 3 Digital Signal

Information can also be represented by a digital signal. For example, a 1 can be encoded as a positive voltage and a 0 as zero voltage. A digital signal can have more than two levels. In this case, we can use zero voltage. A digital signal can have more than two levels. In this case, we can send more than 1 bit for each level. The following figure shows two signals, one with two levels and the other with four. In general, if a signal has L levels, each level needs $\log_2 L$ bits.

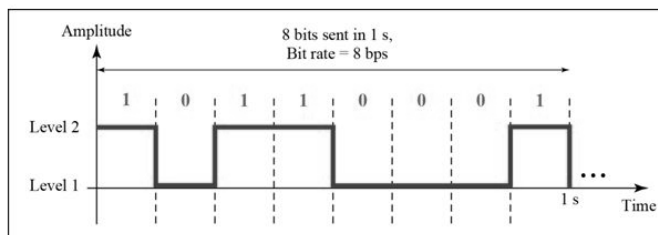


Figure 4 Digital Signal with two levels

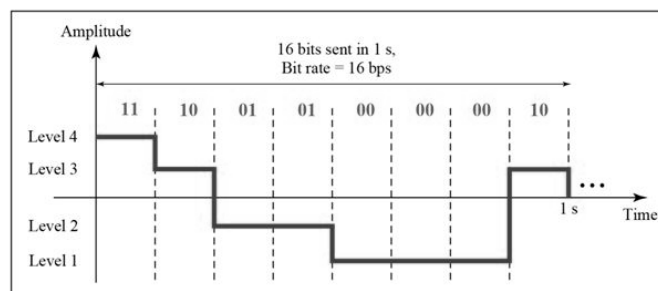


Figure 5 Digital Signal with four levels

Bit Rate:

The bit rate is the number of bits sent in 1s, expressed in bits per second (bps).

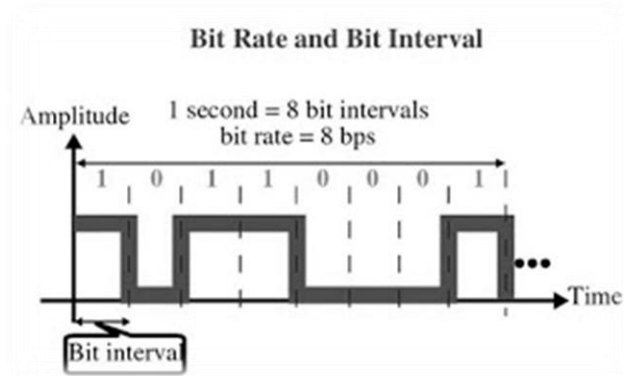


Figure 6 Bit rate and bit interval

The Bit Rate for the above diagram is 8bps and 16bps.

Bit Length:

The bit length is the distance one bit occupies on the transmission medium.

$$\text{Bit length} = \text{propagation speed} \times \text{bit duration}$$

Baud Rate

Baud Rate is the number of signal unit transmitted per second.

Thus Baud Rate is always less than or equal to bit rate. Baud rate is number of symbols per second.

Difference Between Bit Rate and Baud Rate

Bit Rate	Baud Rate
Bit rate is also defined as per second travel number of bits.	Baud rate is also defined as per second number of changes in signal.
Bit rate emphasized on computer efficiency.	While baud rate emphasized on data transmission.
Bit rate is not used to decide the requirement of bandwidth for transmission of signal.	While baud rate is used to decide the requirement of bandwidth for transmission of signal.

Bit Length

The bit length is the distance one bit occupies on the transmission medium.

$$\text{Bit length} = \text{propagation speed} \times \text{bit duration}$$

Bit Interval

Data can be represented by a digital signal. For Example a 1 can be encoded as a positive voltage and a 0 can be encoded as a zero voltage.

2.3 Transmission Mode

Transmission mode refers to the mechanism of transferring of data between two devices connected over a network. It is also called Communication Mode. These modes direct the direction of flow of information.

Modes of Transmission

Transmission mode means transferring of data between two devices. It is also known as communication mode. Buses and networks are designed to allow communication to occur between individual devices that are interconnected. There are three types of transmission mode:-

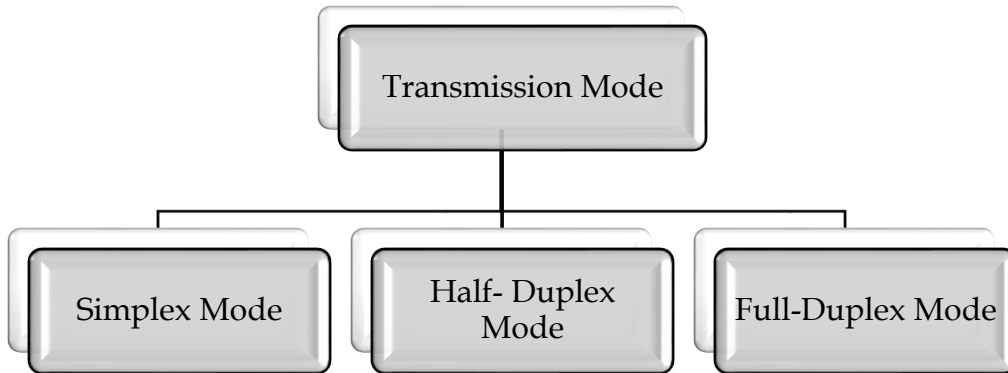


Figure 7 Transmission Modes

These are explained as following below.

a) Simplex Mode -

In Simplex mode, the communication is unidirectional, as on a one-way street. Only one of the two devices on a link can transmit, the other can only receive. The simplex mode can use the entire capacity of the channel to send data in one direction. Example: Keyboard and traditional monitors. The keyboard can only introduce input, the monitor can only give the output.

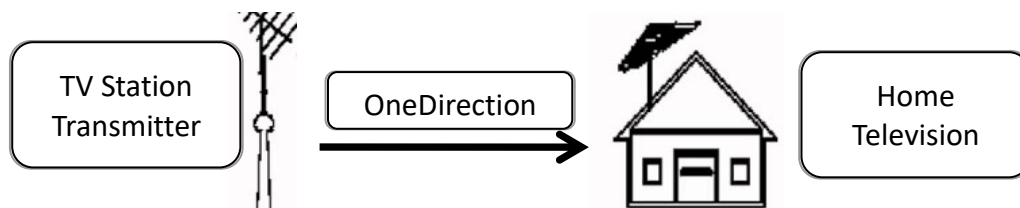


Figure 8 Simplex Mode

b) Half-Duplex Mode -

In half-duplex mode, each station can both transmit and receive, but not at the same time. When one device is sending, the other can only receive, and vice versa. The half-duplex mode is used in cases where there is no need for communication in both direction at the same time. The entire capacity of the channel can be utilized for each direction. Example: Walkie- talkie in which message is sent one at a time and messages are sent in both the directions.

$$\text{Channel capacity} = \text{Bandwidth} * \text{Propagation Delay}$$

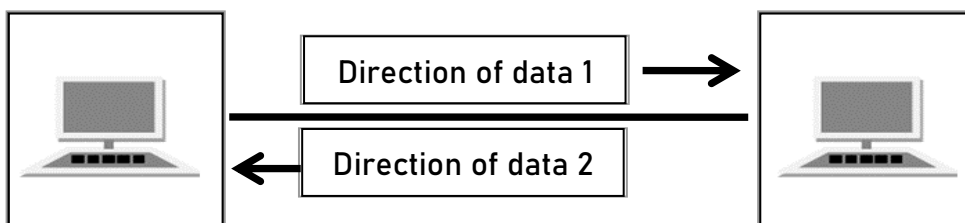


Figure 9 Half Duplex Transmission

c) Full-Duplex Mode -

In full-duplex mode, both stations can transmit and receive simultaneously. In full duplex mode, signals going in one direction share the capacity of the link with signals going in other direction, this sharing can occur in two ways:

Either the link must contain two physically separate transmission paths, one for sending and other for receiving.

Or the capacity is divided between signals travelling in both directions.

Full-duplex mode is used when communication in both direction is required all the time. The capacity of the channel, however, must be divided between the two directions.

Example: Telephone Network in which there is communication between two persons by a telephone line, through which both can talk and listen at the same time.

Channel Capacity = $2 * \text{Bandwidth} * \text{propagation Delay}$

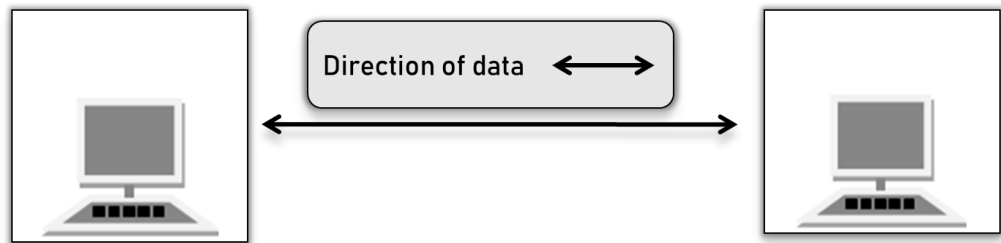


Figure 10 Full Duplex Transmission

Performance Metrics

Network performance can be affected by a number of different factors. It's important for companies to know which network performance metrics are important to examine. However, depending on the specific issues that plague your network, not every metric is going to be important for you to look at. Despite this, there are some metrics that are essential for any businesses to consider.

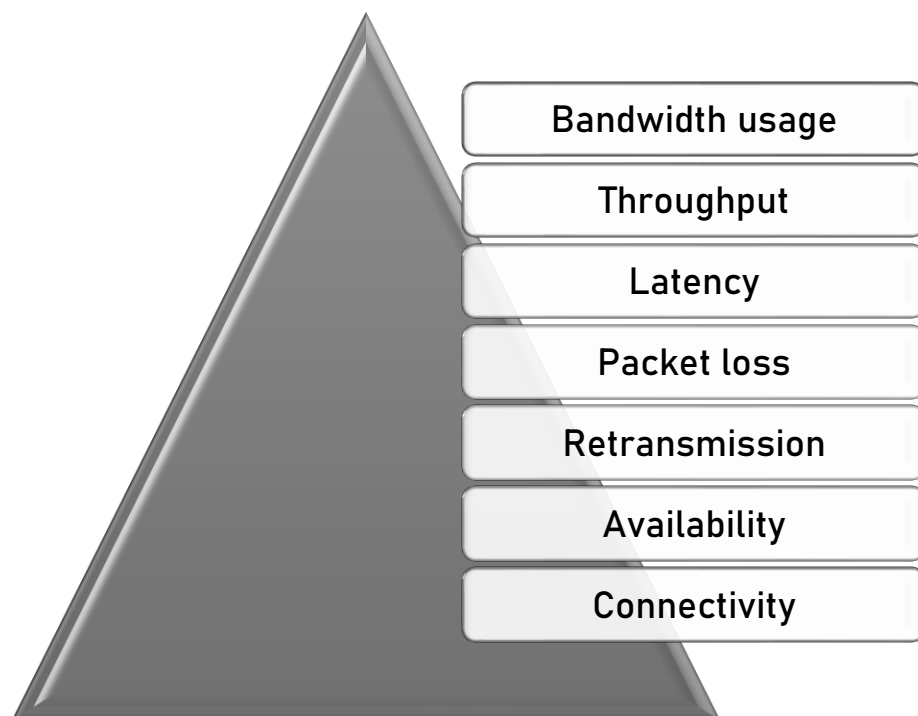


Figure 11: Seven Essential Network Performance Metric Bandwidth usage

Bandwidth is the maximum data transmission rate possible on a network. For optimal network operations, you want to get as close to your maximum bandwidth as possible without reaching critical levels. This indicates that your network is sending as much data as it can within a period of time but isn't being overloaded. An NPM can monitor how much bandwidth is currently being used on a network, as well as how much bandwidth is typically used during daily operations. The solution can also alert you when your network is using too much bandwidth.

Throughput

Throughput measures your network's actual data transmission rate, which can vary wildly through different areas of your network. While your network's bandwidth measures the theoretical limit of data transfer, throughput tells you how much data is actually being sent. Specifically, throughput measures the percentage of data packets that are successfully being sent; a low throughput means there are a lot of failed or dropped packets that need to be sent again.

Latency

Latency is the delay that happens between a node or device requesting data and when that data is finished being delivered. This delay can happen for a variety of reasons, but whatever the cause, your NPM solution can track any delays and log them. Consistent delays or odd spikes in delay time indicate a major performance issue; however, because delays can often be undetectable to the human eye, you need a monitoring tool to keep an eye on any delays that happen.

Packet loss

Packet loss examines how many data packets are dropped during data transmissions on your network. The more data packets that are lost, the longer it takes for a data request to be fulfilled. Your IT team should know how many packets are being dropped on average across your infrastructure. A network's Transmission Control Protocol (TCP) interprets when packets are dropped and takes steps to ensure that data packets can still be transmitted; your network team should monitor this system to make sure it's working.

Retransmission

When packets are lost, the network needs to retransmit it to complete a data request. This retransmission rate lets your enterprise know how often packets are being dropped, which is an indication of congestion on your network. You can analyze retransmission delay, or the time it takes for a dropped packet to be retransmitted, to understand how long it takes your network to recover from packet loss.

Availability

Network availability, also known as uptime, simply measures whether or not the network is currently operational. You can never guarantee 100% availability, but you want to be aware of any downtime that happens on your network that you weren't expecting. It's important to be alerted when the network goes down, which network monitoring tools will provide for you. However, you should also be able to discover your actual uptime percentage and how often your network goes down.

Connectivity

Connectivity refers to whether the connections between the nodes on your network are working properly. If there is an improper or malfunctioning connection on your network, it can be a major hurdle for your company. Ideally, every connection should always be operating at peak levels. However, performance issues like malware can target specific nodes or connections to affect performance in that specific area of the network.

2.4 Transmission Impairments

Signals travel through transmission media, which are not perfect. The imperfection causes signal impairment. This means that the signal at the beginning of the medium is not the same as the signal at the end of the medium. What is sent is not what is received.

Reasons for Impairments

Three causes of impairment are attenuation, distortion, and noise.

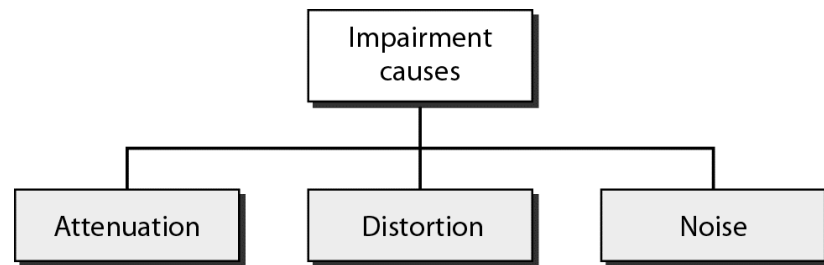


Figure 12 Causes of Impairments

a) Attenuation

It means loss of energy. The strength of signal decreases with increasing distance which causes loss of energy in overcoming resistance of medium. This is also known as attenuated signal. Amplifiers are used to amplify the attenuated signal which gives the original signal back and compensate for this loss.

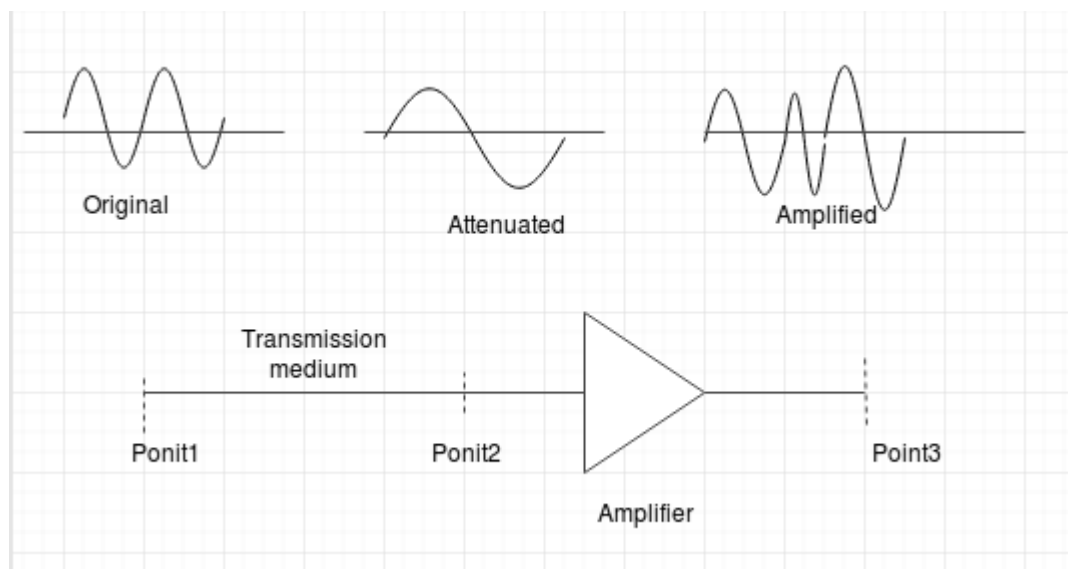


Figure 13 Attenuation

Attenuation is measured in **decibels (dB)**. It measures the relative strengths of two signals or one signal at two different point.

How is Attenuation Measured

To show the loss or gain of energy the unit "decibel" is used.

$$\text{Attenuation dB} = 10 \log_{10} P_2 / P_1$$

P1 - input signal

P2 - output signal

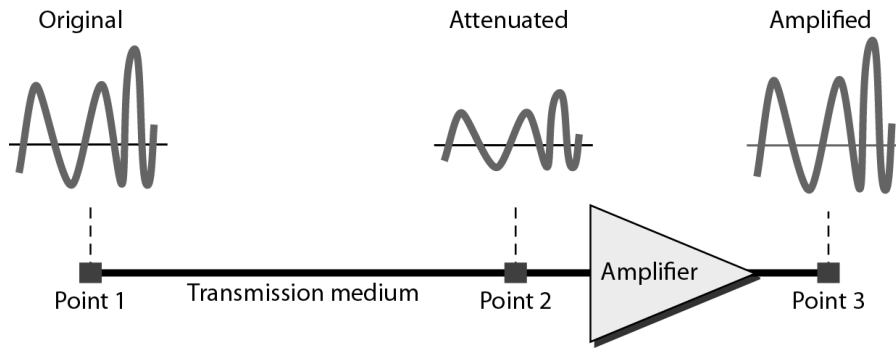


Figure 14 Attenuated Signal

Example 1

Suppose a signal travels through a transmission medium and its power is reduced to one-half. This means that P_2 is $(1/2)P_1$. In this case, the attenuation (loss of power) can be calculated as

$$10 \log_{10} \frac{P_2}{P_1} = 10 \log_{10} \frac{0.5 P_1}{P_1} = 10 \log_{10} 0.5 = 10(-0.3) = -3 \text{ dB}$$

A loss of 3 dB (-3 dB) is equivalent to losing one-half the power.

Example 2

A signal travels through an amplifier, and its power is increased 10 times. This means that $P_2 = 10P_1$. In this case, the amplification (gain of power) can be calculated as

$$10 \log_{10} \frac{P_2}{P_1} = 10 \log_{10} \frac{10P_1}{P_1} = 10 \log_{10} 10 = 10(1) = 10 \text{ dB}$$

Example 3

One reason that engineers use the decibel to measure the changes in the strength of a signal is that decibel numbers can be added (or subtracted) when we are measuring several points (cascading) instead of just two. In Figure 3.27 a signal travels from point 1 to point 4. In this case, the decibel value can be calculated as

$$\text{dB} = -3 + 7 - 3 = +1$$

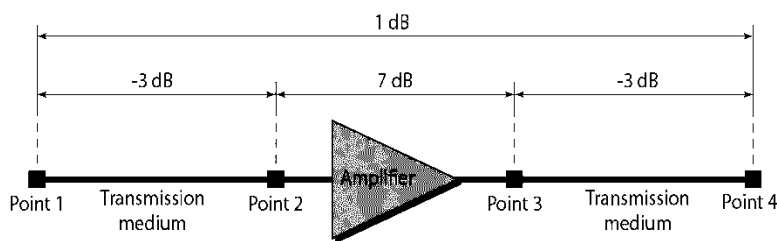


Figure 15 Attenuated Signal over Four Points

b) Distortion

It means changes in the form or shape of the signal. This is generally seen in composite signals made up with different frequencies. Each frequency component has its own propagation speed travelling through a medium. And that's why it delays in arriving at the final destination. Every component arrives at different times which leads to distortion. Therefore, they have different phases at receiver end from what they had at sender's end.

Example

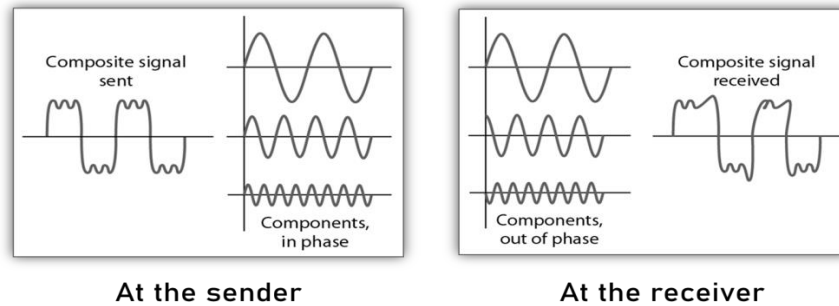


Figure 16 Distortion

c) Noise

The random or unwanted signal that mixes up with the original signal is called noise. There are several types of noise such as induced noise, crosstalk noise, thermal noise and impulse noise which may corrupt the signal.

Induced noise comes from sources such as motors and appliances. These devices act as sending antenna and transmission medium act as receiving antenna. Thermal noise is movement of electrons in wire which creates an extra signal. Crosstalk noise is when one wire affects the other wire. Impulse noise is a signal with high energy that comes from lightning or power lines.

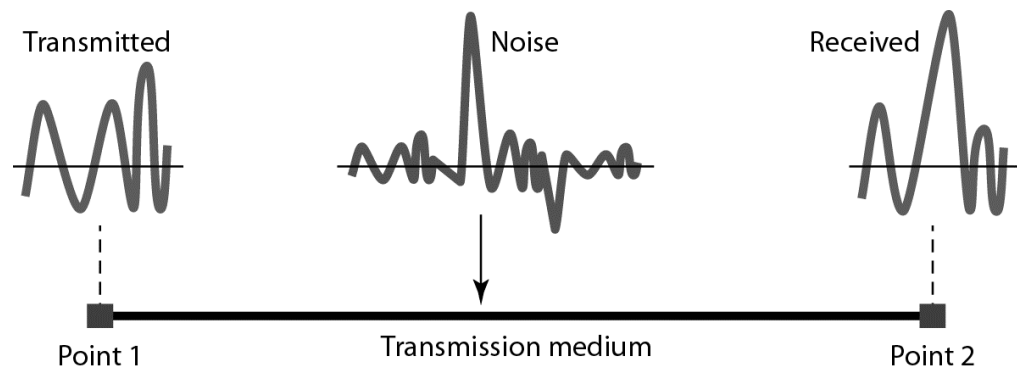


Figure 17 Noise

Signal to Noise Ratio (SNR)

To measure the quality of a system the SNR is often used. It indicates the strength of the signal wrt the noise power in the system.

$$SNR = \frac{\text{Average signal power}}{\text{Average noise power}}$$

It is usually given in dB and referred to as SNR_{dB} .

A high SNR means the signal is less corrupted by noise.

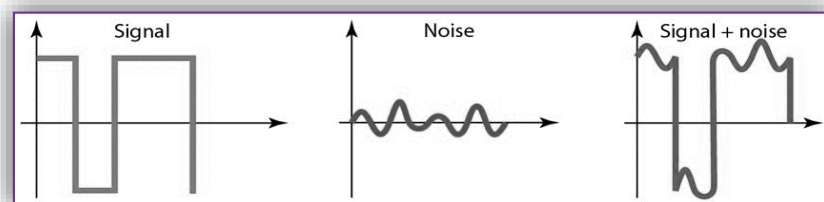


Figure 18 High SNR Signal

- A low SNR means the signal is more corrupted by noise.

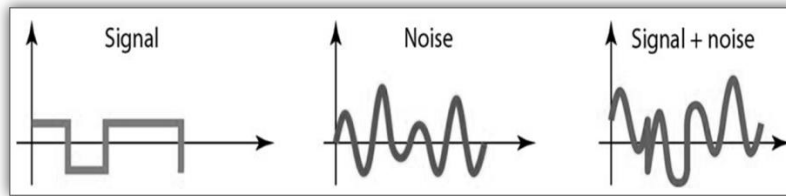


Figure 19 Low SNR Signal

2.5 Protocols

A network protocol is an established set of rules that determine how data is transmitted between different devices in the same network. Essentially, it allows connected devices to communicate with each other, regardless of any differences in their internal processes, structure or design. Cooperative action is necessary. It should be noted that computer networking is not only to exchange bytes. It is a huge system with several utilities and functions. For example error detection, encryption, routing etc. For proper communication, entities in different systems must speak the same language. There must be mutually acceptable conventions and rules about the content, timing and underlying mechanisms. Those conventions and associated rules are referred as "PROTOCOLS".

Protocol Architecture

The task of data transfer is broken up into some modules. It is important to understand as to why it is done and how do these modules interact? For example, file transfer could use three modules: File transfer application, communication service module and network access module. Let us see a real-world example of the Protocol Architecture.

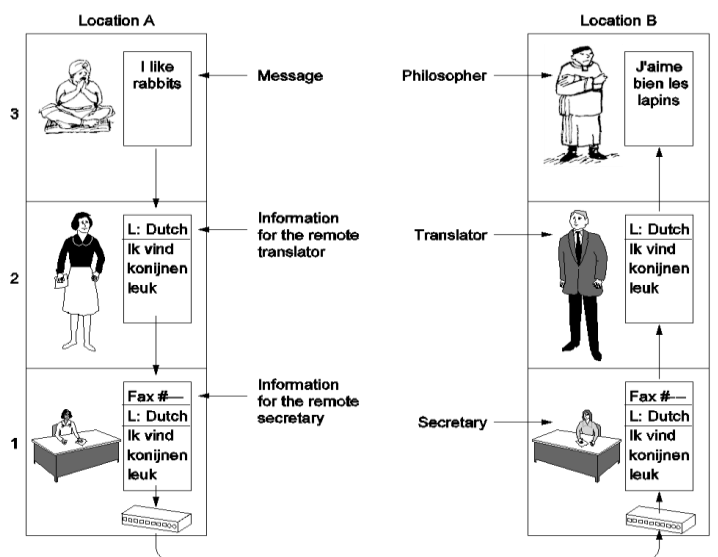


Figure 20 Philosopher-Translator-Secretary Architecture

Let us focus on some of the issues: like the peer-to-peer protocols are independent of each other for example, secretaries may change the comm. medium to email or the translators may agree on using another common language. Note that each layer adds a header

2.6 Network Standards

Networking standards define the rules for data communications that are needed for interoperability of networking technologies and processes. Standards help in creating and maintaining open markets and allow different vendors to compete on the basis of the quality of their products while being compatible with existing market products.

Types of Standards

Standards are of two types i.e. De facto and De jure.

De facto standards are the standards that are followed without any formal plan or approval by any organization. They have come into existence due to traditions or facts. For example, the HTTP had started as a de facto standard.

De jure standards are the standards which have been adopted through legislation by any officially recognized standards organization. Most of the communication standards that are used today are de jure standards.

Standards Organizations

Some of the noted standards organizations are

1. International Standards Organization (ISO)
 2. International Telecommunication Union (ITU)
 3. Institute of Electronics and Electrical Engineers (IEEE)
 4. American National Standards Institute (ANSI)
 5. Internet Research Task Force (IRTF)
 6. Electronic Industries Association (EIA)
 7. World Wide Web Consortium (W3C)
-
1. **International Standards Organization (ISO)**- The International Organization for standardization widely known as ISO, is an international standard-setting body composed of representatives from various national standards organizations. Founded on February 23, 1947, the organization promulgates worldwide proprietary industrial and commercial standards. It has its headquarters in Geneva, Switzerland. While ISO defines itself as a non-governmental organization, its ability to set standards that often-become law, either through treaties or national standards, makes it more powerful than most non-governmental organizations. In practice, ISO acts as a consortium with strong links to governments. ISO is an international standard-setting body composed of representatives from various national standards organizations the organization promulgates worldwide proprietary industrial and commercial standards. ISO's main products are the International Standards. ISO also publishes Technical Reports, Technical Specifications, Publicly Available Specifications, Technical Corrigenda, and Guides.
 2. **International Telecommunication Union (ITU)** - The International Telecommunication Union is the specialized agency of the United Nations which is responsible for information and communication technologies. ITU coordinates the shared global use of the radio spectrum, promotes international cooperation in assigning satellite orbits, works to improve telecommunication infrastructure in the developing world and establishes worldwide standards. ITU also organizes worldwide and regional exhibitions and forums, such as ITU TELECOM WORLD, bringing together representatives of government and the telecommunications and ICT industry to exchange ideas, knowledge and technology.
 3. **Institute of Electronics and Electrical Engineers (IEEE)** -IEEE's Constitution defines the purposes of the organization as "scientific and educational, directed toward the advancement of the theory and practice of Electrical, Electronics, Communications and Computer Engineering, as well as Computer Science, the allied branches of engineering and the related arts and sciences." The IEEE is Not-for-Profit Corporation. It was formed in 1963 by the merger of the Institute of Radio Engineers (IRE, founded 1912) and the American Institute of Electrical Engineers (AIEE, founded 1884). It has more than 400,000 members in more than 160 countries, 45% outside the United States. It is also a leading developer of industrial standards having developed over 900 active industry standards in a broad range of disciplines, including electric power and energy, biomedical technology and health care, information technology, information assurance, telecommunications, consumer electronics, transportation, aerospace,

and nanotechnology. IEEE develops and participates in educational activities such as accreditation of electrical engineering programs in institutes of higher learning. IEEE serves as a major publisher of scientific journals and a conference organizer. IEEE is one of the leading standards-making organizations in the world. IEEE performs its standards making and maintaining functions through the IEEE Standards Association (IEEE-SA). One of the more notable IEEE standards is the IEEE 802 LAN/MAN group of standards which includes the IEEE 802.3 Ethernet standard and the IEEE 802.11 Wireless Networking standard.

4. **American National Standards Institute (ANSI)** -Though ANSI itself does not develop standards, the Institute oversees the development and use of standards by accrediting the procedures of standards developing organizations. ANSI accreditation signifies that the procedures used by standards developing organizations meet the Institute's requirements for openness, balance, consensus, and due process. ANSI was originally formed in 1918, when five engineering societies and three government agencies founded the American Engineering Standards Committee (AESC). In 1928, the AESC became the American Standards Association (ASA). In 1966, the ASA was reorganized and became the United States of America Standards Institute (USASI). The present name was adopted in 1969. ANSI also designates specific standards as American National Standards, or ANS, when the Institute determines that the standards were developed in an environment that is equitable, accessible and responsive to the requirements of various stakeholders. The American National Standards process involves:
 - consensus by a group that is open to representatives from all interested parties
 - broad-based public review and comment on draft standards
 - consideration of and response to comments
 - incorporation of submitted changes that meet the same consensus requirements into a draft standard
 - availability of an appeal by any participant alleging that these principles were not respected during the standards-development process.
5. **Internet Research Task Force (IRTF)** -The Internet Research Task Force (IRTF) promotes research of importance to the evolution of the Internet by creating focused, long-term Research Groups working on topics related to Internet protocols, applications, architecture, and technology. The IRTF is composed of several focused and long-term Research Groups. Research Groups have the stable long-term membership needed to promote the development of research collaboration and teamwork in exploring research issues. Participation is by individual contributors, rather than by representatives of organizations. The IRTF is managed by the IRTF Chair in consultation with the Internet Research Steering Group (IRSG). The IRSG membership includes the IRTF Chair, the chairs of the various Research Groups, and other individuals ("members at large") from the research community selected by the IRTF Chair. The IRTF is managed by the IRTF Chair in consultation with the Internet Research Steering Group (IRSG). The IRSG membership includes the IRTF Chair, the chairs of the various Research Groups, and other individuals ("members at large") from the research community selected by the IRTF Chair.
6. **World Wide Web Consortium (W3C)** -The World Wide Web Consortium (W3C) is the main international standards organization for World Wide Web (abbreviated WWW or W3). Founded and headed by Tim Berners-Lee, the consortium is made up of member organizations which maintain full-time staff for the purpose of working together in the development of standards for the World Wide Web. W3C also engages in education and outreach, develops software and serves as an open forum for discussion about the Web. W3C

was created to ensure compatibility and agreement among industry members in the adoption of new standards. Prior to its creation, incompatible versions of HTML were offered by different vendors, increasing the potential for inconsistency between web pages. The consortium was created to get all those vendors to agree on a set of core principles and components which would be supported by everyone.

Summary

Having discussed the various transmission medias and the different type of signals that they deal with, a proper transmission mode has to be chosen for the type of communication we are undergoing. To ensure the effective and error-free transmission different essential network performance metrics must be considered to make necessary changes in the network. The different transmission impairments need to be understood by an organisation and should be effectively managed to ensure effective data communication

Keywords

Archive: A computer site advertises and stores a large amount of public domain, shareware software and documentation.

Broadcast Networks: They have a single communication channel, which is shared by all the computers on the network and therefore, any message transmitted by a computer on the network is received by all the computers connected to the channel.

Bit Length: It is the distance one bit occupies on the transmission medium.

Baud Rate is the number of signal unit transmitted per second. It is always less than or equal to bit rate.

Distortion: It means changes in the form or shape of the signal. This is generally seen in composite signals made up with different frequencies. Each frequency component has its own propagation speed travelling through a medium. And thats why it delays in arriving at the final destination

Self-Assessment

Fill in the blanks

1.refers to whether the connections between the nodes on your network are working properly.
2.measures your network's actual data transmission rate, which can vary wildly through different areas of your network.
3.is the delay that happens between a node or device requesting data and when that data is finished being delivered.
4.examines how many data packets are dropped during data transmissions on your network.
5.rate lets your enterprise know how often packets are being dropped, which is an indication of congestion on your network.

State whether the following is true or false.

6. Bandwidth is the minimum data transmission rate possible on a network.
7. De facto standards are the standards that are followed with a formal plan or approval by any organization.
8. De jure standards are the standards which have been adopted through legislation by any officially recognized standards organization.
9. The World Wide Web Consortium (W3C) is the main international standards organization for World Wide Web which was founded and headed by Tim Berners-Lee.
10. SNR measures the quality of a system that indicates the strength of the signal wrt the noise power in the system.

Answers for Self Assessment

1. Connectivity 2. Throughput 3. Latency 4. Packet loss 5. Retransmission
6. False 7. False 8. True 9. True 10. True

Review Questions

1. Underline the key differences between Bit Rate and Baud Rate. Also elaborate on Bit length and Bit Interval.
2. Signals can be classified based on different parameters. Elaborate the different classification categories.
3. Explain the different factors that can affect a Network Performance. Explain the metrics that are essential for any businesses to consider.
4. Explain how the imperfections in the transmission medias causes signal impairments. Explain the various types of impairments.
5. Compare and contrast the various types of transmission modes. Take suitable examples to explain.

Further Readings



Andrew S. Tanenbaum, *Computer Networks*, Prentice Hall

Behrouz A. Forouzan and Sophia Chung Fegan, *Data Communications and Networking*, McGraw-Hill Companies

Burton, Bill, *Remote Access for Cisco Networks*, McGraw-Hill Osborne Media McGraw-Hill Osborne Media

Rajneesh Agrawal and Bharat Bhushan Tiwari, *Computer Networks and Internet*, Vikas Publication

Unit 03: Network Models

CONTENTS

Objectives

Introduction

3.1 The Layered Architecture

3.2 Basic Elements of Layered Architecture

3.3 Need of Layered Architecture?

3.4 Network Models

3.5 Functions of the OSI Layers

3.6 Transmission Control Protocol (TCP)

3.7 Internet Addresses

3.8 Layered Tasks

Summary

Keywords

Self Assessment

Answer for Self Assessment

Further Readings

Objectives

After this lecture, you would be able to

- learn about the standard layer protocol suites adopted for data transmission in computer networks.
- learn about the standard layer protocol suites adopted for data transmission in computer networks.
- learn about the standard layer protocol suites adopted for data transmission in computer networks.
- Understand the architecture of TCP/IP protocol suite
- Differentiate between the OSI model and TCP/IP Suite.
- Differentiate between three types of internet addresses.

Introduction

The field of computer networks has evolved through the highly structured development phases. The main task of the system designer is to scale up and upgrade the networks does reducing the design complexity with the help of a structured Network Architecture and structure. This involves standardizing the rules or protocols for layering the communication process. The main objective is to break down the communication process into smaller and easy to handle independent categories which salt some distinct aspects of the data exchange process. Thus, the layers on one computer converse with the corresponding layers on the other computer in the network. These rules and conventions used are collectively known as layer protocols.

3.1 The Layered Architecture

For data communication to take place and two or more users can transmit data from one to other, a systematic approach is required. This approach enables users to communicate and transmit data

through efficient and ordered path. It is implemented using models in computer networks and are known as computer network models. Computer network models are responsible for establishing a connection among the sender and receiver and transmitting the data in a smooth manner respectively.

The main aim of the layered architecture is to divide the design into small pieces. Each lower layer adds its services to the higher layer to provide a full set of services to manage communications and run the applications. It provides modularity and clear interfaces, i.e., provides interaction between subsystems. It ensures the independence between layers by providing the services from lower to higher layer without defining how the services are implemented. Therefore, any modification in a layer will not affect the other layers. The number of layers, functions, contents of each layer will vary from network to network. However, the purpose of each layer is to provide the service from lower to a higher layer and hiding the details from the layers of how the services are implemented.

3.2 Basic Elements of Layered Architecture

The basic elements of layered architecture are services, protocols, and interfaces.

1. **Service:** It is a set of actions that a layer provides to the higher layer.
2. **Protocol:** It defines a set of rules that a layer uses to exchange the information with peer entity. These rules mainly concern about both the contents and order of the messages used.
3. **Interface:** It is a way through which the message is transferred from one layer to another layer.

In a layer-n architecture, layer-n on one machine will have a communication with the layer-n on another machine and the rules used in a conversation are known as a layer-n protocol. In case of layered architecture, no data is transferred from layer-n of one machine to layer-n of another machine. Instead, each layer passes the data to the layer immediately just below it, until the lowest layer is reached. Layer 1 is the physical medium through which the actual communication takes place. In a layered architecture, unmanageable tasks are divided into several small and manageable tasks. The data is passed from the upper layer to lower layer through an interface. A layered architecture as shown in Table 3.1 Functions of OSI Model provides a clean-cut interface so that minimum information is shared among different layers. It also ensures that the implementation of one layer can be easily replaced by another implementation. A set of layers and protocols is known as network architecture.

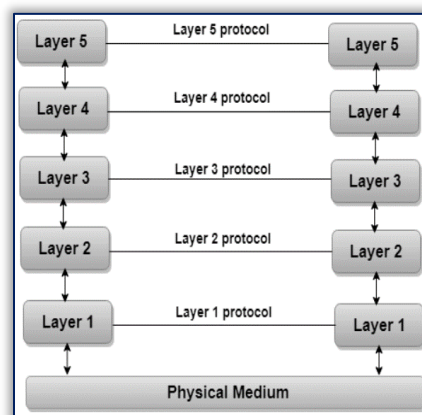


Figure 3.1 The Layered Architecture

3.3 Need of Layered Architecture?

The main task of the system designer is to scale up and upgrade the networks does reducing the design complexity with the help of a structured Network Architecture and structure. This involves standardizing the rules or protocols for layering the communication process. The main objective is to break down the communication process into smaller and easy to handle independent categories which salt some distinct aspects of the data exchange process. Thus the layered on one computer converse with the corresponding layers on the other computer in the network. These rules and

conventions used are collectively known as layer protocols. The main features of the layered approach are:

- Divide-and-Conquer Approach:** The divide-and-conquer approach makes a design process in such a way that the unmanageable tasks are divided into small and manageable tasks. In short, we can say that this approach reduces the complexity of the design.
- Modularity:** Layered architecture is more modular. Modularity provides the independence of layers, which is easier to understand and implement.
- Easy to modify:** It ensures the independence of layers so that implementation in one layer can be changed without affecting other layers.
- Easy to test:** Each layer of the layered architecture can be analysed and tested individually.

3.4 Network Models

There are two computer network models (as shown in Figure 3.2) on which the whole data communication process relies.

- The OSI Reference Model
- The TCP/IP Model

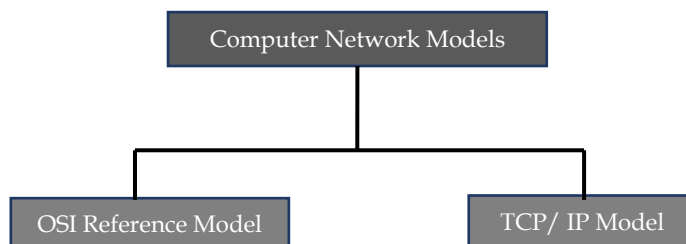


Figure 3.2 Computer Network Models

1. The OSI Model

Established in 1947, the International Standards Organization (ISO) is a multinational body dedicated to worldwide agreement on international standards. An ISO standard that covers all aspects of network communications is the Open Systems Interconnection (OSI) model. It was first introduced in the late 1970s.

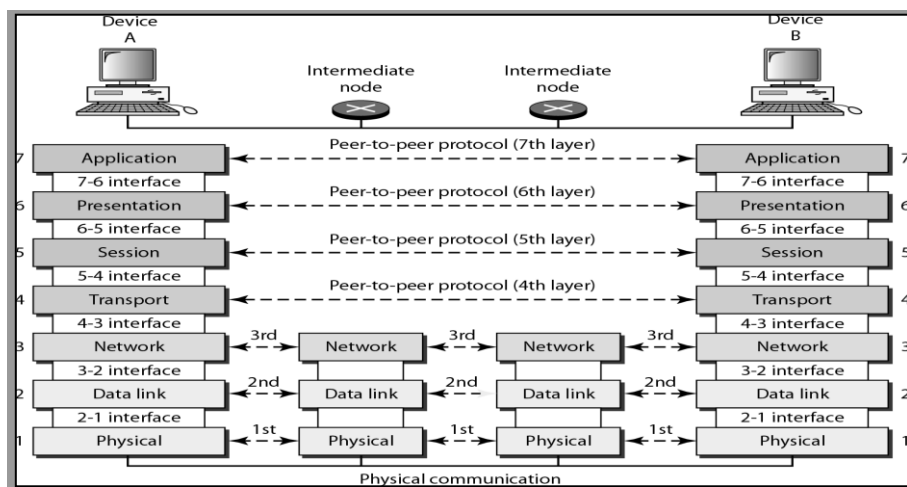


Figure 3.3 Layer wise view of the OSI model

OSI stands for Open System Interconnection is a reference model (as shown in Figure 3.3) that describes how information from a software application in one computer moves through a physical medium to the software application in another computer. OSI consists of seven layers, and each layer performs a particular network function. OSI model was developed by the International Organization for Standardization (ISO) in 1984, and it is now considered as an architectural model

for the inter-computer communications. OSI model divides the whole task into seven smaller and manageable tasks. Each layer is assigned a particular task. Each layer is self-contained, so that task assigned to each layer can be performed independently.

The OSI model is divided into two layers:

1. Upper layers
2. Lower layers.

The upper layer of the OSI model mainly deals with the application related issues, and they are implemented only in the software. The application layer is closest to the end user. Both the end user and the application layer interact with the software applications. The lower layer of the OSI model deals with the data transport issues. The data link layer and the physical layer are implemented in hardware and software. The physical layer is the lowest layer of the OSI model and is closest to the physical medium. The physical layer is mainly responsible for placing the information on the physical medium. The layer categories of OSI model can be seen in Figure 3.4.

The OSI Reference Model (Layer Categorization)

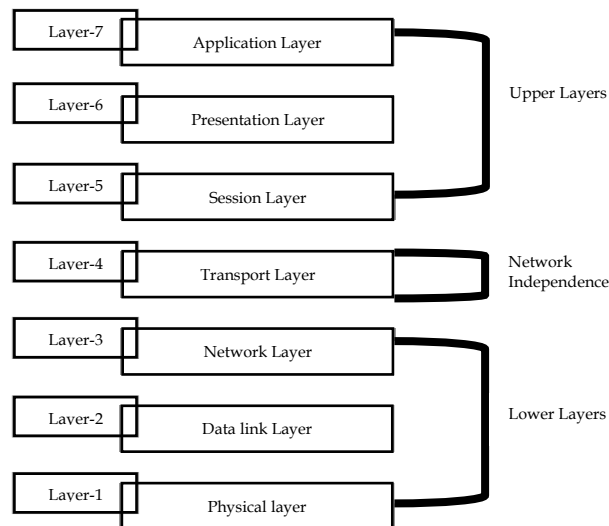


Figure 3.4 Layered Categorization of OSI Model

Functions of the OSI Layers

The functions of the OSI model are as demarcated in Table 3.1

Table 3.1 Functions of OSI Model

Layer No.	Layer	Responsibility of	PDU (Protocol Data Unit)	Data Flow	Address Used
7	Application Layer	Host	Data		Specific Address
6	Presentation Layer	Host	Data		Specific Address
5	Session Layer	Host	Data		Specific Address
4	Transport Layer	Host	Segment	Process to Process	Port Address
3	Network Layer	Network	Packet	End to End	Logical Address
2	Data Link Layer	Network	Frame	Hop to Hop	Physical Address
1	Physical Layer	Network	Bits and Bytes		

Encapsulation & De-encapsulation in OSI Model can be understood with the help of Figure 3.5.

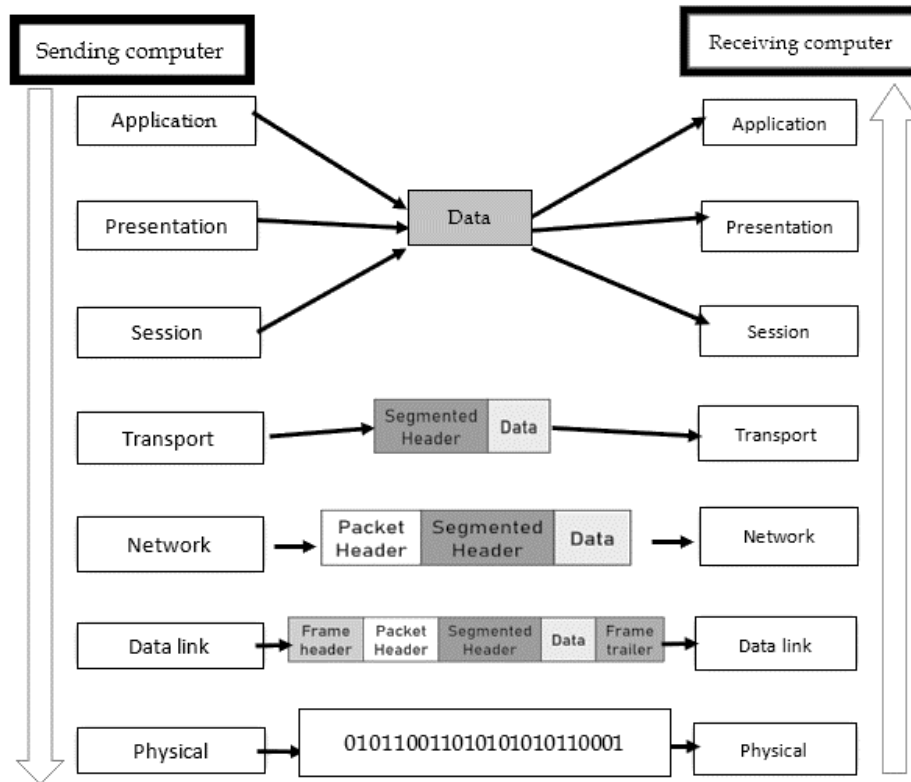


Figure 3.5 Encapsulation & De-encapsulation in OSI Model

3.5 Functions of the OSI Layers

The seven layers of the OSI model performs the different essential functions

Application Layer:

This layer is responsible for providing interface to the application user. This layer encompasses protocols which directly interact with the user. It serves as a window for users and application processes to access network service. It handles issues such as network transparency, resource allocation, etc. An application layer is not an application, but it performs the application layer functions. This layer provides the network services to the end-users as shown in Figure 3.6.

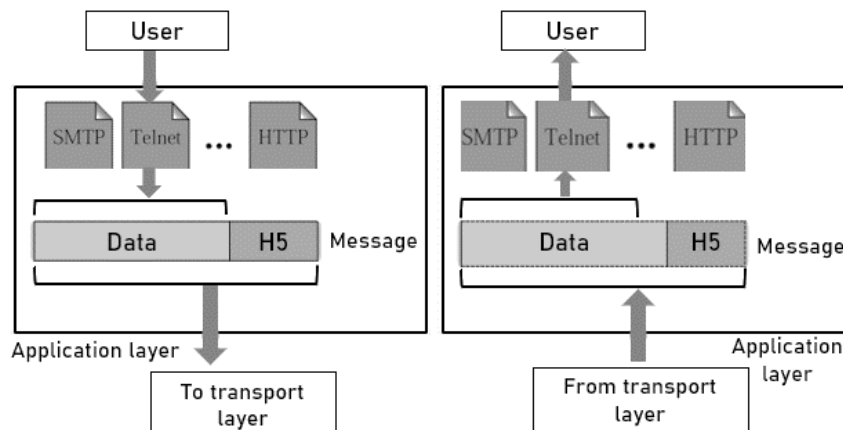


Figure 3.6 Services Provided by Application Layer

Functions of Application layer:

The application layer performs the following vital functions as shown in Figure 3.7. It is used for communicating among applications. Its main task is of supporting network applications like ftp, smtp, http etc. An application layer is the topmost layer in the TCP/IP model. It is responsible for handling high-level protocols, issues of representation. This layer allows the user to interact with the application. When one application layer protocol wants to communicate with another

application layer, it forwards its data to the transport layer. There is an ambiguity occurs in the application layer. Every application cannot be placed inside the application layer except those who interact with the communication system. For example: text editor cannot be considered in application layer while web browser using **HTTP** protocol to interact with the network where **HTTP** protocol is an application layer protocol.

- **File transfer, access, and management (FTAM):** An application layer allows a user to access the files in a remote computer, to retrieve the files from a computer and to manage the files in a remote computer.
- **Mail services:** An application layer provides the facility for email forwarding and storage.
- **Directory services:** An application provides the distributed database sources and is used to provide that global information about various objects.

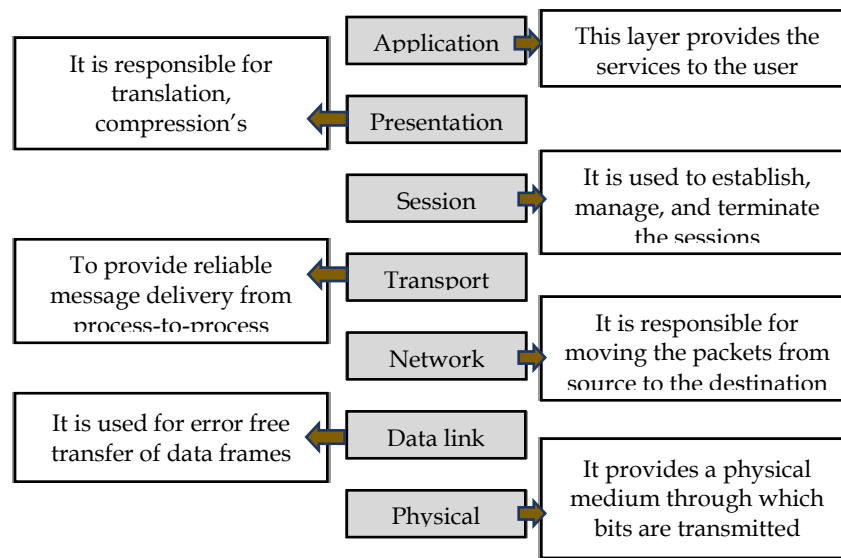


Figure 3. 7 Functions of Application layer

- **HTTP:** HTTP stands for Hypertext transfer protocol. This protocol allows us to access the data over the world wide web. It transfers the data in the form of plain text, audio, video. It is known as a Hypertext transfer protocol as it has the efficiency to use in a hypertext environment where there are rapid jumps from one document to another.
- **SNMP:** SNMP stands for Simple Network Management Protocol. It is a framework used for managing the devices on the internet by using the TCP/IP protocol suite.
- **SMTP:** SMTP stands for Simple mail transfer protocol. The TCP/IP protocol that supports the e-mail is known as a Simple mail transfer protocol. This protocol is used to send the data to another e-mail address.
- **DNS:** DNS stands for Domain Name System. An IP address is used to identify the connection of a host to the internet uniquely. But, people prefer to use the names instead of addresses. Therefore, the system that maps the name to the address is known as Domain Name System.
- **TELNET:** It is an abbreviation for Terminal Network. It establishes the connection between the local computer and remote computer in such a way that the local terminal appears to be a terminal at the remote system.
- **FTP:** FTP stands for File Transfer Protocol. FTP is a standard internet protocol used for transmitting the files from one computer to another computer.

Presentation Layer:

The Presentation layer defines how data in the native format of remote host should be presented in the native format of host. It is mainly concerned with the syntax and semantics of the information exchanged between the two systems. It acts as a data translator for a network. This layer is a part of

the operating system that converts the data from one presentation format to another format. The Presentation layer is also known as the syntax layer. The main services performed by presentation layer can be understood as shown in Figure 3.8.

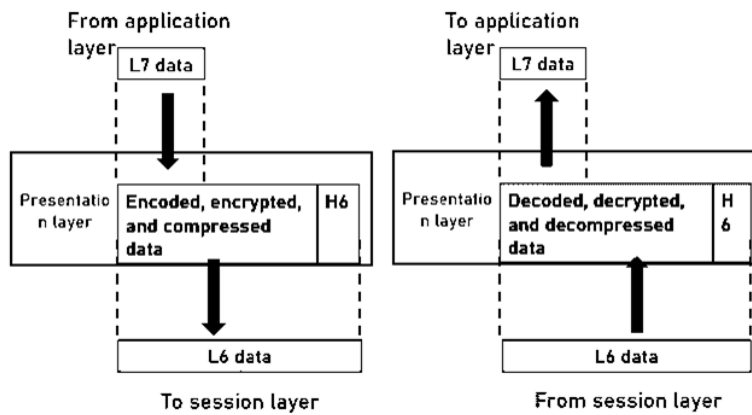


Figure 3.8 Services Provided by Presentation Layer

- **Translation:** The processes in two systems exchange the information in the form of character strings, numbers and so on. Different computers use different encoding methods, the presentation layer handles the interoperability between the different encoding methods. It converts the data from sender-dependent format into a common format and changes the common format into receiver-dependent format at the receiving end.
- **Encryption:** Encryption is needed to maintain privacy. Encryption is a process of converting the sender-transmitted information into another form and sends the resulting message over the network.
- **Compression:** Data compression is a process of compressing the data, i.e., it reduces the number of bits to be transmitted. Data compression is very important in multimedia such as text, audio, video.
- **Session Layer:** This layer maintains sessions between remote hosts. For example, once user/password authentication is done, the remote host maintains this session for a while and does not ask for authentication again in that time span. It is a layer 5 in the OSI model. The Session layer is used to establish, maintain and synchronizes the interaction between communicating devices as shown in Figure 3.9.

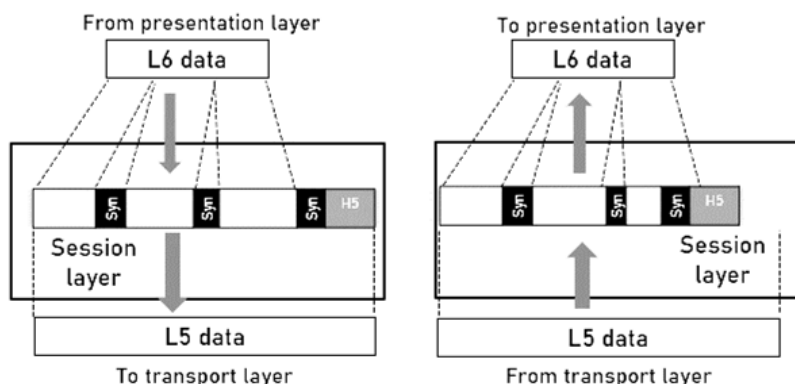


Figure 3.9 Tasks Performed by Session Layer

- **Dialog control:** Session layer acts as a dialog controller that creates a dialog between two processes, or we can say that it allows the communication between two processes which can be either half-duplex or full duplex.
- **Synchronization:** Session layer adds some checkpoints when transmitting the data in a sequence. If some error occurs in the middle of the transmission of data, then the transmission

will take place again from the checkpoint. This process is known as Synchronization and recovery.

Transport Layer

This layer is responsible for process-to-process delivery between hosts. It is the Layer 4 ensures that messages are transmitted in the order in which they are sent and there is no duplication of data. The main responsibility of the transport layer is to transfer the data completely. It receives the data from the upper layer and converts them into smaller units known as segments. This layer can be termed as a process to process layer as it provides a point-to-point connection between source and destination to deliver the data reliably as shown in Figure 3.10.

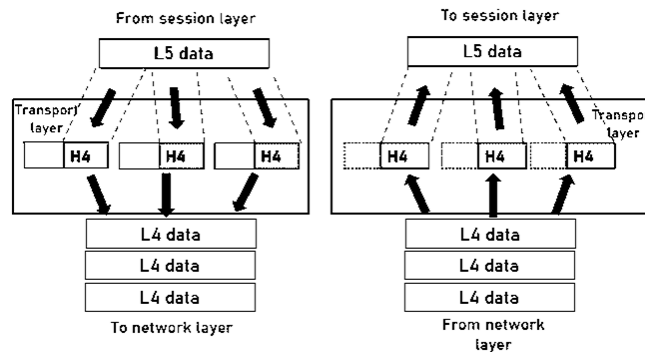


Figure 3. 10 Services Provided by Transport Layer

The Two Protocols Used in Transport Layer

1. Transmission Control Protocol
2. User Datagram Protocol

1. Transmission Control Protocol

It is a standard protocol that allows the systems to communicate over the internet. It establishes and maintains a connection between hosts. When data is sent over the TCP connection, then the TCP protocol divides the data into smaller units known as segments. Each segment travels over the internet using multiple routes, and they arrive in different orders at the destination. The transmission control protocol reorders the packets in the correct order at the receiving end.

2. User Datagram Protocol

User Datagram Protocol is a transport layer protocol. It is an unreliable transport protocol as in this case receiver does not send any acknowledgment when the packet is received, the sender does not wait for any acknowledgment. Therefore, this makes a protocol unreliable.

Functions of Transport Layer

The various functions performed by the transport layer are:

- **Service-point addressing:** Computers run several programs simultaneously due to this reason, the transmission of data from source to the destination not only from one computer to another computer but also from one process to another process. The transport layer adds the header that contains the address known as a service-point address or port address. The responsibility of the network layer is to transmit the data from one computer to another computer and the responsibility of the transport layer is to transmit the message to the correct process.
- **Segmentation and reassembly:** When the transport layer receives the message from the upper layer, it divides the message into multiple segments, and each segment is assigned with a sequence number that uniquely identifies each segment. When the message has arrived at the destination, then the transport layer reassembles the message based on their sequence numbers.

- **Connection Control:** Transport layer provides two services Connection-oriented service and connectionless service. A connectionless service treats each segment as an individual packet, and they all travel in different routes to reach the destination. A connection-oriented service makes a connection with the transport layer at the destination machine before delivering the packets. In connection-oriented service, all the packets travel in the single route.
- **Flow control:** The transport layer also responsible for flow control but it is performed end-to-end rather than across a single link.
- **Error control:** The transport layer is also responsible for Error control. Error control is performed end-to-end rather than across the single link. The sender transport layer ensures that message reach at the destination without any error.

Network Layer

The Network layer is responsible for address assignment and uniquely addressing hosts in a network. It is a layer 3 that manages device addressing, tracks the location of devices on the network. It determines the best path to move data from source to the destination based on the network conditions, the priority of service, and other factors.

Functions of Network Layer

The various functions of Network Layer areas shown in Figure 3.11:

- **Internetworking:** An internetworking is the main responsibility of the network layer. It provides a logical connection between different devices.
- **Addressing:** A Network layer adds the source and destination address to the header of the frame. Addressing is used to identify the device on the internet.
- **Routing:** Routing is the major component of the network layer, and it determines the best optimal path out of the multiple paths from source to the destination.
- **Packetizing:** A Network Layer receives the segments from the upper layer and converts them into packets. This process is known as Packetizing. It is achieved by internet protocol (IP).

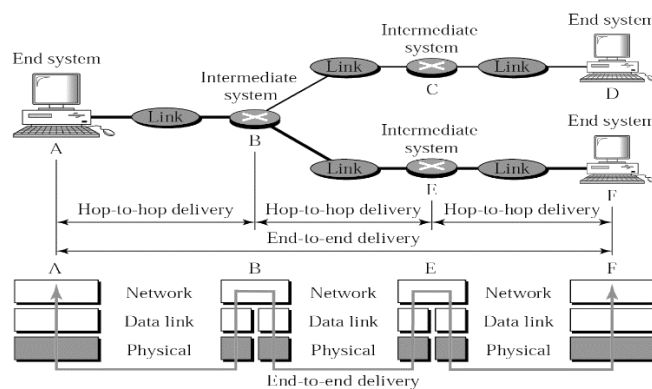


Figure 3. 11 Packetization Process

Data Link Layer

The **Data Link Layer** is responsible for routing and forwarding the packets. Routers are the layer 3 devices, they are specified in this layer and used to provide the routing services within an internetwork. The protocols used to route the network traffic are known as Network layer protocols. Examples of protocols are IP and Ipv6.

This layer is responsible for reading and writing data from and onto the line. Link errors are detected at this layer. This layer is responsible for the error-free transfer of data frames. It defines the format of the data on the network. It provides a reliable and efficient communication between two or more devices. It is mainly responsible for the unique identification of each device that resides on a local network.

It contains two sub-layers:

1. Logical Link Control Layer
2. Media Access Control Layer

1. Logical Link Control Layer

- It is responsible for transferring the packets to the Network layer of the receiver that is receiving.
- It identifies the address of the network layer protocol from the header.
- It also provides flow control.

2. Media Access Control Layer

- A Media access control layer is a link between the Logical Link Control layer and the network's physical layer.
- It is used for transferring the packets over the network.

Functions of the Data-link layer

Framing: The data link layer translates the physical's raw bit stream into packets known as Frames. The Data link layer adds the header and trailer to the frame. The header which is added to the frame contains the hardware destination and source address.

- **Physical Addressing:** The Data link layer adds a header to the frame that contains a destination address. The frame is transmitted to the destination address mentioned in the header.
- **Flow Control:** Flow control is the main functionality of the Data-link layer. It is the technique through which the constant data rate is maintained on both the sides so that no data get corrupted. It ensures that the transmitting station such as a server with higher processing speed does not exceed the receiving station, with lower processing speed.
- **Error Control:** Error control is achieved by adding a calculated value CRC (Cyclic Redundancy Check) that is placed to the Data link layer's trailer which is added to the message frame before it is sent to the physical layer. If any error seems to occur, then the receiver sends the acknowledgment for the retransmission of the corrupted frames.
- **Access Control:** When two or more devices are connected to the same communication channel, then the data link layer protocols are used to determine which device has control over the link at a given time. It can be seen in Figure 3.12.

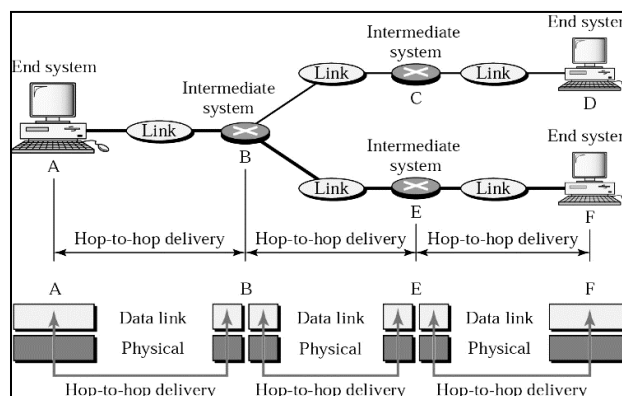


Figure 3. 12 Access Control Functionality

Physical Layer

Physical layer defines the hardware, cabling wiring, power output, pulse rate etc. The main functionality of the physical layer is to transmit the individual bits from one node to another node as shown in Figure 3.13. It is the lowest layer of the OSI model. It establishes, maintains

and deactivates the physical connection. It specifies the mechanical, electrical and procedural network interface specifications.

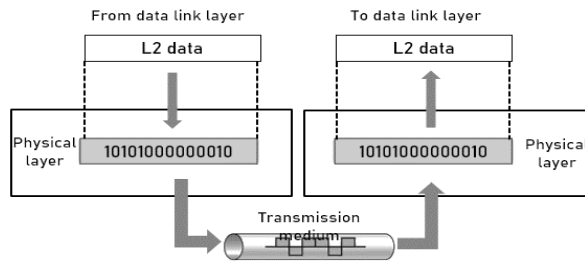


Figure 3. 13 Working of Physical Layer

Functions of Physical Layer

- **Line Configuration:** It defines the way how two or more devices can be connected physically.
- **Data Transmission:** It defines the transmission mode whether it is simplex, half-duplex or full-duplex mode between the two devices on the network.
- **Topology:** It defines the way how network devices are arranged.
- **Signals:** It determines the type of the signal used for transmitting the information.

The various functions can also be understood with the help of Figure 3.14.

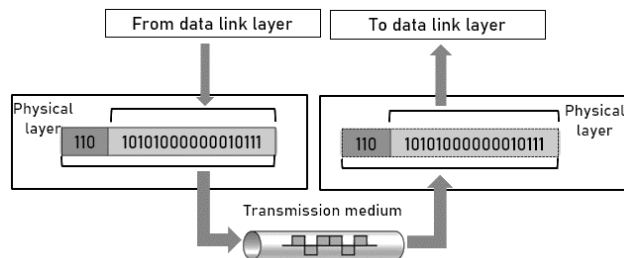


Figure 3. 14 Functions Performed by Physical Layer

Transport Layer:

The Transport layer is responsible for providing reliable delivery of data. It provides host-host data transfer. The main protocols are TCP and UDP.

Internet Layer:

It defines uniform format of packets forwarded across networks of different technologies and rules for forwarding packets in routers

Network Access Layer:

Defines formats for carrying packets in hardware frames. It is Responsible for routing of a datagram from source to destination. It takes care of the IP and routing protocols. It makes use of Link for data transfer between neighboring network elements. It makes use of Point-to-Point Protocol (PPP) and the Ethernet. It handles the responsibility of transferring physical bits “on the wire”.

Network Layer

A network layer is the lowest layer of the TCP/IP model. It is the combination of the Physical layer and Data Link layer defined in the OSI reference model. It defines how the data should be sent physically through the network. This layer is mainly responsible for the transmission of the data between two devices on the same network. The functions carried out by this layer are encapsulating the IP datagram into frames transmitted by the network and mapping of IP addresses into physical addresses. The protocols used by this layer are ethernet, token ring, FDDI, X.25, frame relay.

Internet Protocol (IP)

IP protocol is used in this layer, and it is the most significant part of the entire TCP/IP suite. The responsibilities of this protocol are:

- **IP Addressing:** This protocol implements logical host addresses known as IP addresses. The IP addresses are used by the internet and higher layers to identify the device and to provide internetwork routing.
- **Host-to-host communication:** It determines the path through which the data is to be transmitted.

Internet Layer

An internet layer is also known as the network layer. The main responsibility of the internet layer is to send the packets from any network, and they arrive at the destination irrespective of the route they take. Following protocols are used at this layer:

1. Internet Protocol (IP)
2. Address Resolution Protocol (ARP)
3. Internet Control Message Protocol (ICMP)

1. Internet Protocol (IP)

- **Data Encapsulation and Formatting:** An IP protocol accepts the data from the transport layer protocol. An IP protocol ensures that the data is sent and received securely, it encapsulates the data into message known as IP datagram.
- **Fragmentation and Reassembly:** The limit imposed on the size of the IP datagram by data link layer protocol is known as Maximum Transmission unit (MTU). If the size of IP datagram is greater than the MTU unit, then the IP protocol splits the datagram into smaller units so that they can travel over the local network. Fragmentation can be done by the sender or intermediate router. At the receiver side, all the fragments are reassembled to form an original message.
- **Routing:** When IP datagram is sent over the same local network such as LAN, MAN, WAN, it is known as direct delivery. When source and destination are on the distant network, then the IP datagram is sent indirectly. This can be accomplished by routing the IP datagram through various devices such as routers.

2. Address Resolution Protocol (ARP)

ARP stands for **Address Resolution Protocol**. It is a network layer protocol which is used to find the physical address from the IP address. The two terms are mainly associated with the ARP Protocol:

a. ARP request

b. ARP reply

a. **ARP request:** When a sender wants to know the physical address of the device, it broadcasts the ARP request to the network.

b. **ARP reply:** Every device attached to the network will accept the ARP request and process the request, but only recipient recognize the IP address and sends back its physical address in the form of ARP reply. The recipient adds the physical address both to its cache memory and to the datagram header

3. ICMP Protocol

ICMP stands for Internet Control Message Protocol. It is a mechanism used by the hosts or routers to send notifications regarding datagram problems back to the sender. A datagram travels from router-to-router until it reaches its destination. If a router is unable to route the data because of some unusual conditions such as disabled links, a device is on fire or network congestion, then the ICMP protocol is used to inform the sender that the datagram is undeliverable.

An ICMP protocol mainly uses two terms:

- **ICMP Test:** ICMP Test is used to test whether the destination is reachable or not.
- **ICMP Reply:** ICMP Reply is used to check whether destination device is responding or not.

The core responsibility of the ICMP protocol is to report the problems, not correct them. The responsibility of the correction lies with the sender. ICMP can send the messages only to the source, but not to the intermediate routers because the IP datagram carries the addresses of the source and destination but not of the router that it is passed to.

User Datagram Protocol (UDP)

Provides connectionless service and end-to-end delivery of transmission. It is an unreliable protocol as it discovers the errors but not specify the error. It discovers the error, and ICMP protocol reports the error to the sender that user datagram has been damaged. UDP does not specify which packet is lost. UDP contains only checksum; it does not contain any ID of a data segment.

- UDP consists of the following fields:
 - **Source port address:** The source port address is the address of the application program that has created the message.
 - **Destination port address:** The destination port address is the address of the application program that receives the message.
 - **Total length:** It defines the total number of bytes of the user datagram in bytes.
 - **Checksum:** The checksum is a 16-bit field used in error detection.

Summary of Layers

We use the concept of layers in our daily life. As an example, let us consider two friends who communicate through postal mail. The process of sending a letter to a friend would be complex if there were no services available from the post office.

The functions performed by the different layers can be summarized as in Figure 3.23

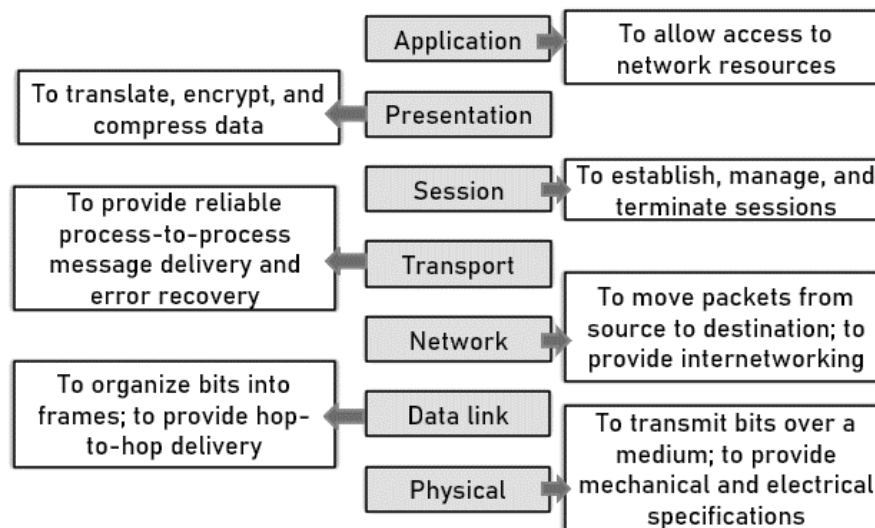


Figure 3. 23 Functions Performed by Different Layers of OSI Model

3.6 Transmission Control Protocol (TCP)

TCP/IP Protocol Suite

The original TCP/IP protocol suite was defined as having four layers: host-to-network, internet, transport, and application.

However, in comparison to OSI, we can say that the TCP/IP protocol suite is made of five layers: physical, data link, network, transport, and application.

The layers in the TCP/IP protocol suite do not exactly match those in the OSI model. The original TCP/IP protocol suite was defined as having four layers: host-to-network, internet, transport, and application. However, when TCP/IP is compared to OSI, we can say that the TCP/IP protocol suite is made of five layers: physical, data link, network, transport, and application.

Layered Tasks

The different layers are:

- Application
- Transport
- Internet
- Network Interface
- Physical

It provides a full transport layer services to applications. It creates a virtual circuit between the sender and receiver, and it is active for the duration of the transmission. TCP is a reliable protocol as it detects the error and retransmits the damaged frames. Therefore, it ensures all the segments must be received and acknowledged before the transmission is considered to be completed and a virtual circuit is discarded. At the sending end, TCP divides the whole message into smaller units known as segment, and each segment contains a sequence number which is required for reordering the frames to form an original message. At the receiving end, TCP collects all the segments and reorders them based on sequence numbers.

TCP/IP and OSI model

OSI model is a reference model while TCP/IP is an implementation of OSI model as in Figure 3.24.

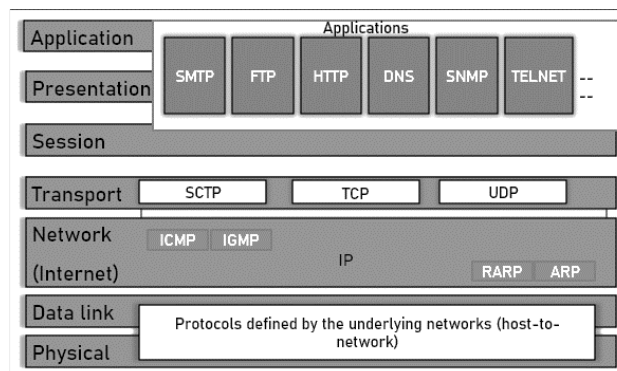


Figure 3.144 OSI vs TCP/IP Model

- All the layers from bottom till Transport Layer provides End to End Transport Service
- All the layers above the Transport Layer are application oriented and use the Transport Service.

The differences between OSI and the TCP/IP model can be summarized with the Table 3.2

Table 3.2 Functions of OSI Model

OSI Model	TCP/IP Model
The OSI model however is a "generic, protocol-independent standard"	TCP/IP Protocols are considered to be standards around which the internet has developed.
OSI having presentation layer and session layer independently	TCP/IP combines the presentation and session layer issues into its application layer.
OSI having data link and physical layers independently	TCP/IP combines the OSI data link and physical layers into the network access layer.
OSI appears to complex because of 7 layers	TCP/IP appears to be a more simpler model and this is mainly due to the fact that it has fewer layers.

Networks are not usually built around the OSI model as it is merely used as a guidance tool.	TCP/IP is considered to be a more credible model- This is mainly due to the fact because TCP/IP protocols are the standards around which the internet was developed therefore it mainly gains creditability due to this reason
The OSI model is bottom to up process of network connection	TCP/IP is the top to bottom process structure for internet purpose.
OSI defines several more layers of standardized functions	TCP/IP makes no assumptions about what happens above the level of a network session
OSI model is a reference model	TCP/IP is an implementation of OSI model.
OSI provides layer functioning and also defines functions of all the layers.	TCP/IP model is more based on protocols and protocols are not flexible with other layers.
In OSI model the transport layer guarantees the delivery of packets	In TCP/IP model the transport layer does not guarantees delivery of packets.
Follows horizontal approach	Follows vertical approach.
OSI model defines services, interfaces and protocols very clearly and makes clear distinction between them	In TCP/IP it is not clearly separated its services, interfaces and protocols
OSI is a general model	TCP/IP model cannot be used in any other application.
Network layer of OSI model provide both connection oriented and connectionless service.	The Network layer in TCP/IP model provides connectionless service.
It has 7 layers	It has 4 layers
Transport layer guarantees delivery of packets	Transport layer does not guarantees delivery of packets
The protocol are better hidden and can be easily replaced as the technology changes.	It is not easy to replace the protocols
OSI truly is a general model	TCP/IP can not be used for any other application
OSI mode represents an idea	TCP/IP network model represents reality in the world

OSI	TCP/IP
Application	Application
Presentation	
Session	
Transport	Transport
Network	Network
Data link	Physical
Physical	

Figure 3.25 The Functions of OSI as Performed by different layers of TCP/IP

Similarities between OSI and TCP/IP

The similarities between OSI and TCP/IP model can be graphically understood with Figure 3.25

- They share similar architecture.
- They share a common application layer.
- Knowledge by networking professionals.
- Both models assume that packets are switched.
- Both the reference models are based upon layered architecture.
- The physical layer and the data link layer of the OSI model correspond to the link layer of the TCP/IP model.
- The network layers and the transport layers are the same in both the models.
- The session layer, the presentation layer and the application layer of the OSI model together form the application layer of the TCP/IP model.
- In both the models, protocols are defined in a layer-wise manner.
- In both models, data is divided into packets and each packet may take the individual route from the source to the destination.

3.7 Internet Addresses

- Four levels of addresses are used in an internet employing the TCP/IP protocols: physical, logical, port, and specific.
- Physical Address
- Logical Address
- Port Address
- Specific address

Addressing

The different types of addresses can be understood with the help of Figure 3.26

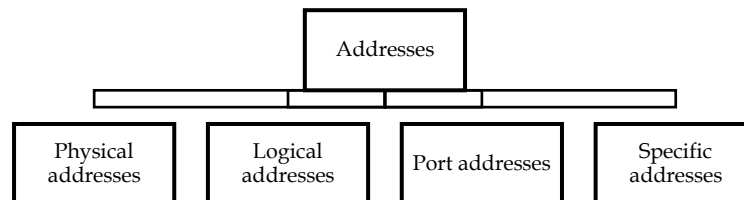
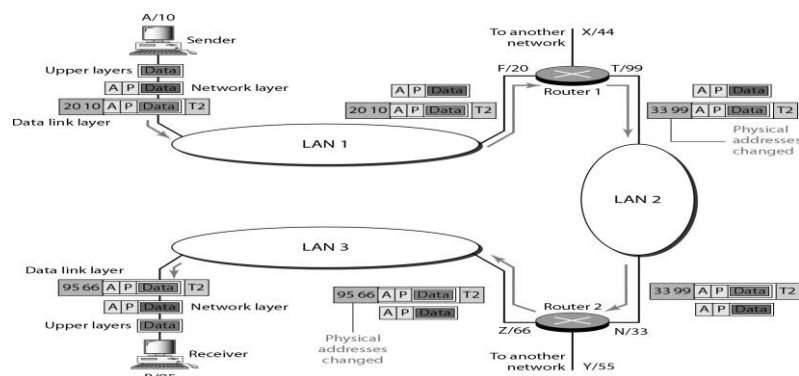


Figure 3.26 Different types of Addresses

Logical Address

It shows a part of an internet with two routers connecting three LANs. Each device (computer or router) has a pair of addresses (logical and physical) for each connection. In this case, each computer is connected to only one link and therefore has only one pair of addresses. Each router, however, is connected to three networks (only two are shown in the figure 3.30). So each router has three pairs of addresses, one for each connection.



Port Address

Figure 3. 15 Logical Addresses for Routers

It shows two computers communicating via the Internet. The sending computer is running three processes currently with port addresses a, b, and c. The receiving computer is running two processes at this time with port addresses j and k. Process a in the sending computer needs to communicate with process j in the receiving computer. Note that although physical addresses change from hop to hop, logical and port addresses remain the same from the source to destination.



The physical addresses will change from hop to hop, but the logical addresses usually remain the same.



753

A 16-bit port address represented as one single number



The physical addresses change from hop to hop, but the logical and port addresses usually remain the same.

A port address is a 16-bit address represented by one decimal number as shown.

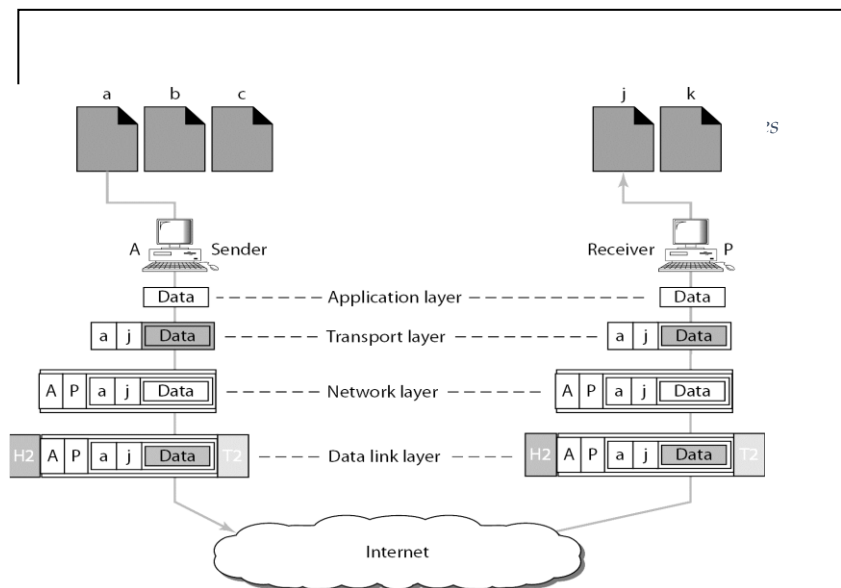
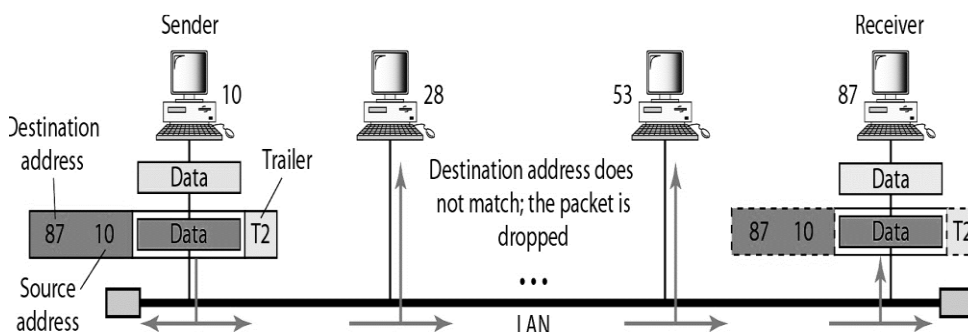


Figure 3. 17 A Network Scenario



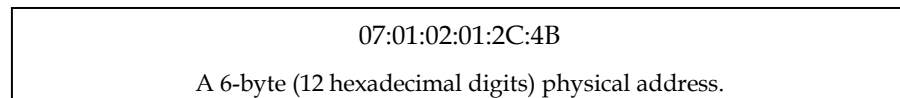
Example

In a node with physical address 10 sends a frame to a node with physical address 87. The two nodes are connected by a link (bus topology LAN). As the figure shows, the computer with physical address 10 is the sender, and the computer with physical address 87 is the receiver. See Figure 3.29.





Most local-area networks use a 48-bit (6-byte) physical address written as 12 hexadecimal digits; every byte (2 hexadecimal digits) is separated by a colon, as shown below:



Relationship of Layers and Addresses in TCP/IP

The relationship of layers and addresses in TCP/IP model can be seen in Figure 3.27.

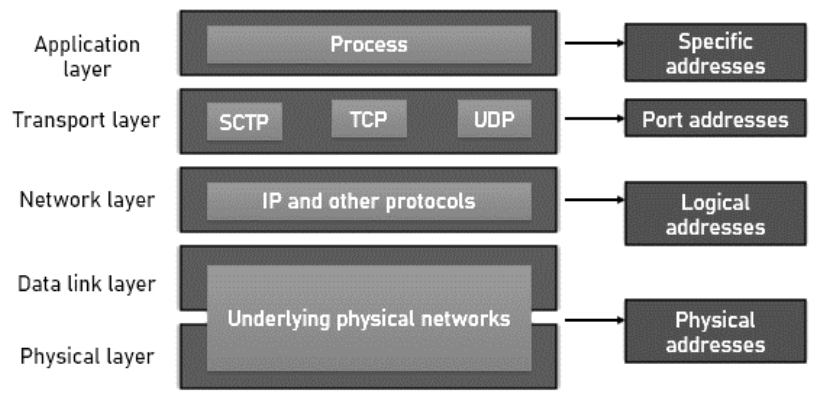


Figure 3. 27 Relationship of Layers and Addresses in TCP/IP

Advantages of OSI Model

The OSI Model being one of the most widely used computer network models does possess some major advantages which makes it so popular.

- Each layer has its definite structure and functionality which makes OSI model simple and easy to use.
- It is a general-purpose reference model that can be used for data communication.
- Connection oriented and connection-less services are supported.
- Connection between any type of devices or host or hardware or software is possible.

Disadvantages of OSI Model

- Because of its inability to fit protocols, this model was replaced by TCP/IP Internet Model.
- Session and Presentation layers does not provide high end functionalities and are not of much use as compared to other layers.
- Connection oriented and connection-less services are supported.
- Connection between any type of devices or host or hardware or software is possible.

3.8 Layered Tasks

We use the concept of layers in our daily life. As an example, let us consider two friends who communicate through postal mails shown in Figure 3.28. The process of sending a letter to a friend would be complex if there were no services available from the post office.

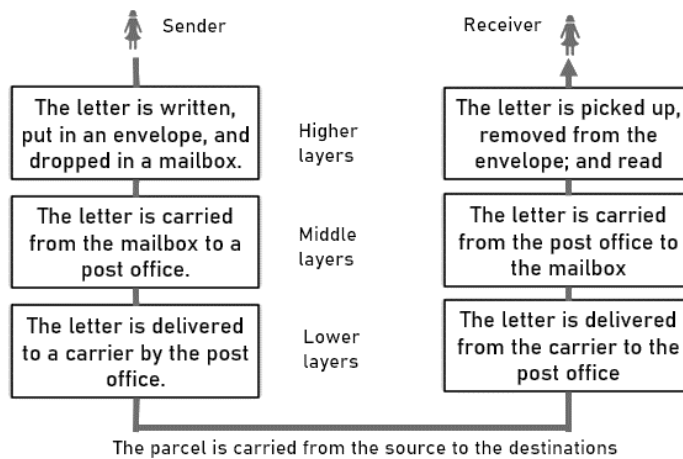


Figure 3. 18Tasks Involved in Sending a Letter

Summary

The three basic components namely, hardware, protocols (software) and applications (useful software) are mandatory to implement a computer network. It is also explained that the concept of layers is important in networking. Each layer with two layers works as the interface and protects the upper layer that each one layer can change with minimum impact on the upper layers. In some cases, this protection is so proficient that an application may not know that it is running on different hardware. The OSI network model has seven layers. TCP/IP stands for Transmission Control Protocol/Internet Protocol. It was developed with the objective to specify a suite of protocols capable of providing transparent communications interoperability services between computers of all sizes, regardless of the hardware or operating system platforms supporting them. Over the years, TCP/IP has become the most widespread of today's protocols. One reason for TCP/IP's popularity is the public availability of its protocols' specifications. In this sense, TCP/IP can justifiably be considered an open system. Most users rely on TCP/IP for the purpose of file transfers, electronic mail (e-mail), and remote login services.

Keywords

Internet Protocol: The Internet protocol suite is the set of communications protocols used for the Internet and other similar networks.

Open Systems Interconnection (OSI) Reference Model: The International Standardization Organization (ISO) developed the OSI model of data communications in 1984. OSI specifies an even-layer model that is used by the industry as the frame of reference when describing protocol architectures and functional characteristics.

TCP/IP: Transmission Control Protocol (TCP) and Internet Protocol (IP) are two distinct network protocols, technically speaking. TCP and IP are so commonly used together; however, that TCP/IP has become standard terminology to refer to either or both of the protocols.

Self Assessment

State whether the following statements are true or false:

1. The entities comprising the corresponding layers on different computers are called clients.
2. The International Organization for Standardization (ISO) took the initiative in setting up OSI.
3. Data communication process allocates memory resources, commonly known as communications buffers for the sake of transmission and reception of data.
4. The information exchanged between two computers is physically carried by means of chemical signals assuming certain coding methods.

5. OSI reference model divides the required functions of the network architecture into five layers and defines the function of each layer.
6. The TCP/IP layer corresponds to the network layer of the OSI reference model in functionality.
7. TCP is a protocol and UDP is an unreliable connectionless protocol.
8. OSI reference model divides the required functions of the into several layers and defines the function of each layer.
9. are entities in the same layer on different computers.
10. is the point from where services can be accessed?
11. What layer in the TCP/IP stack is equivalent to the Transport layer of the OSI model?
 - A. Application
 - B. Host to host
 - C. Internet
 - D. Network Access
12. Which of the following protocols uses both TCP and UDP?
 - A. FTP
 - B. SMTP
 - C. Telnet
 - D. DNS
13. TCP/IP layer is equivalent to combined Session, Presentation and _____
 - A. Network layer
 - B. Application layer
 - C. Transport layer
 - D. Physical layer
14. How many levels of addressing is provided in TCP/IP protocol?
 - A. One
 - B. Two
 - C. Three
 - D. Four
15. A device operating at network layer is called _____
 - A. Router
 - B. Equalizer
 - C. Bridge
 - D. Repeater

Answer for Self Assessment

- | | | | | |
|-------------|---------------------------------|-------------------------|------------------|---------------------------|
| 1. False | 2. True | 3. True | 4. False | 5. False |
| 6. internet | 7. Reliable connection-oriented | 8. Network architecture | 9. Peer entities | 10. Service access points |
| 11. B | 12. D | 13. B | 14. D | 15. A |

Review Questions

1. What are the important design issues for the information exchange among computers?
2. What are the major functions of the network layer in the ISO-OSI model? How the function of packet delivery of network layer is different from data link layer?
3. What is the purpose of layer isolation in the OSI reference model?
4. Why OSI Reference model was widely adopted? What did it make to set itself as a standard for data communication?
5. Highlight the differences between OSI reference model and TCP/IP model.

Further Readings



Andrew S. Tanenbaum, *Computer Networks*, Prentice Hall

Behrouz A. Forouzan and Sophia Chung Fegan, *Data Communications and Networking*, McGraw-Hill Companies

Burton, Bill, *Remote Access for Cisco Networks*, McGraw-Hill Osborne Media McGraw-Hill Osborne Media

Rajneesh Agrawal and Bharat Bhushan Tiwari, *Computer Networks and Internet*, Vikas Publication



<https://www.geeksforgeeks.org/basics-computer-networking/>

Unit 04: Physical Layer

CONTENTS

Objectives

Introduction

4.1 Understanding the Physical Layer

4.2 Functions of Physical Layer

4.3 What is Transmission Media?

4.4 Propagation Modes

4.5 Out of phase signal

4.6 Introduction to Networking Devices

Summary

Keywords

Self Assessment

Review Questions

Answers: Self Assessment

Further Readings

Objectives

After this lecture, you would be able to

- Understand the various services provided by the Physical layer
- Learn about the wired and the wireless transmission medias
- learn about the various networking devices.
- understand the working and functionality of various networking devices.

Introduction

In today's information age the main concern his conveyance for transmission of valuable information across distances using some form of transmission media. Information can be transmitted from the source to the destination in the form of electrical signals. For interactive communication selection of proper medium please a very vital role. For transmitting signal there are different between different guided and unguided medias. The guided or the wired media provides various choices between metal conductors like Twisted pair and Coaxial cable and non-metal conductors such as optical fibers. The choices are governed by various factors such as attenuation, bandwidth, cost of the network etc. The important point to be understood is that during transmission of signal the data is encoded to energy and then energy is transmitted. Similarly, at the receiving end the energy is converted back to the data. This energy can be electrical light, radio waves etc. This energy can be electrical, light and radio energy etc. This transmitted energy can be best transmitted through a compatible transmission medium. This may involve use of special hardware for data encoding and connection to transmission medium. So, the most optimal bounded or unbounded media is selected for transmission.

4.1 Understanding the Physical Layer

Well! The Physical layer is the lowest layer of the OSI reference model. It is responsible for sending bits from one computer to another. This layer is not concerned with the meaning of the bits and deals with the setup of physical connection to the network and with transmission and reception of signals.

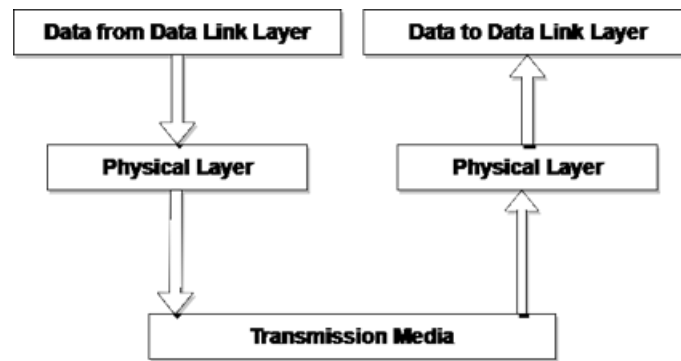


Figure 4. 1 Position of Physical Layer

4.2 Functions of Physical Layer

Representation of Bits: Data in this layer consists of stream of bits. The bits must be encoded into signals for transmission. It defines the type of encoding i.e. how 0's and 1's are changed to signal.

- **Data Rate:** This layer defines the rate of transmission which is the number of bits per second.
- **Synchronization:** It deals with the synchronization of the transmitter and receiver. The sender and receiver are synchronized at bit level.
- **Interface:** The physical layer defines the transmission interface between devices and transmission medium.
- **Line Configuration:** This layer connects devices with the medium: Point to Point configuration and Multipoint configuration.
- **Topologies:** Devices must be connected using the following topologies: Mesh, Star, Ring and Bus.
- **Transmission Modes:** Physical Layer defines the direction of transmission between two devices: Simplex, Half Duplex, Full Duplex.
 - Deals with baseband and broadband transmission

4.3 What is Transmission Media?

Transmission media is a communication channel that carries the information from the sender to the receiver. Data is transmitted in the form of an electromagnetic signals. The main functionality of the transmission media is to carry the information in the form of bits through LAN(Local Area Network). It is a physical path between transmitter and receiver in data communication. In a copper-based network, the bits in the form of electrical signals. In a fibre-based network, the bits travel in the form of light pulses. The transmission media is available in the lowest layer of the OSI reference model, i.e., **Physical layer**. In **OSI or the** (Open System Interconnection) model, transmission media supports the Layer 1. Therefore, it is considered to be as a Layer 1 component. The electrical signals can be sent through the copper wire, fibre optics, atmosphere, water, and vacuum. The characteristics and quality of data transmission are determined by the characteristics of medium and signal. Transmission media is of two types, which are the wired and the wireless media. In wired media, medium characteristics are more important whereas, in wireless media, signal characteristics are more important. Different transmission media have different properties such as bandwidth, delay, cost and ease of installation and maintenance.

Factors affecting Transmission Medias

The various factors affecting Transmission Medias are

- **Bandwidth:** All the factors are remaining constant, the greater the bandwidth of a medium, the higher the data transmission rate of a signal.

- **Transmission impairment:** When the received signal is not identical to the transmitted one due to the transmission impairment. The quality of the signals will get destroyed due to transmission impairment.

Causes of Transmission Impairment

- **Attenuation:** Attenuation means the loss of energy, i.e., the strength of the signal decreases with increasing the distance which causes the loss of energy.
- **Distortion:** Distortion occurs when there is a change in the shape of the signal. This type of distortion is examined from different signals having different frequencies. Each frequency component has its own propagation speed, so they reach at a different time which leads to the delay distortion.
- **Noise:** When data is travelled over a transmission medium, some unwanted signal is added to it which creates the noise.
- **Interference:** An interference is defined as the process of disrupting a signal when it travels over a communication medium on the addition of some unwanted signal.

Classification of Transmission Media

Transmission media is divided into two classes:

1. *Guided (wired) Media*
2. *Unguided (wireless) Media*

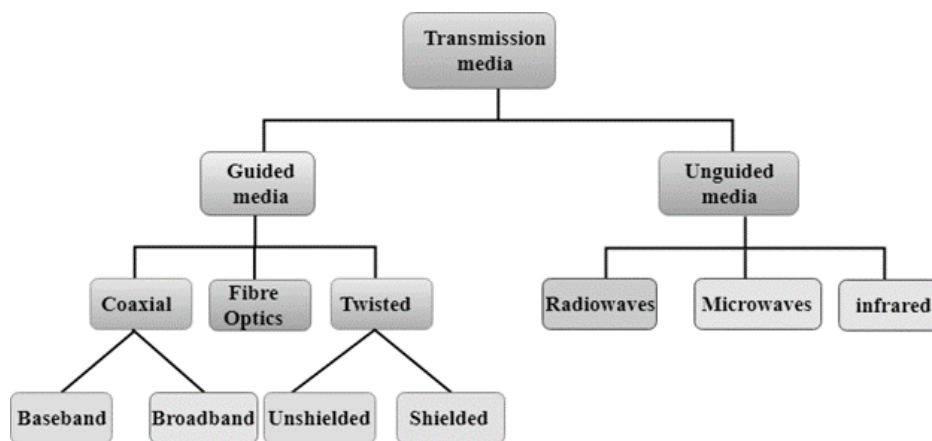


Figure 4. 2 Classification of Transmission Media

1. Guided Media

It is defined as the physical medium through which the signals are transmitted. It is also known as Bounded Media. Now let us discuss the various types Of Guided media:

Twisted Pair	Coaxial Cable	Fibre Optic
• Unshielded Twisted Pair • Shielded Twisted Pair	• Baseband • Broadband	• Single Mode • Multimode • Step Index • Graded Index

Figure 4. 3 Transmission Medias and Types

a) Twisted Pair

Twisted pair is a physical media made up of a pair of cables twisted with each other. A twisted pair cable is cheap as compared to other transmission media. Installation of the twisted pair cable is easy, and it is a lightweight cable. The frequency range for twisted pair cable is from 0 to 3.5KHz.



Figure 4. 4 Structure of Twisted Pair

A twisted pair consists of two insulated copper wires arranged in a regular spiral pattern. The degree of reduction in noise interference is determined by the number of turns per foot. Increasing the number of turns per foot decreases noise interference or the attenuation.

Types of Twisted Pair

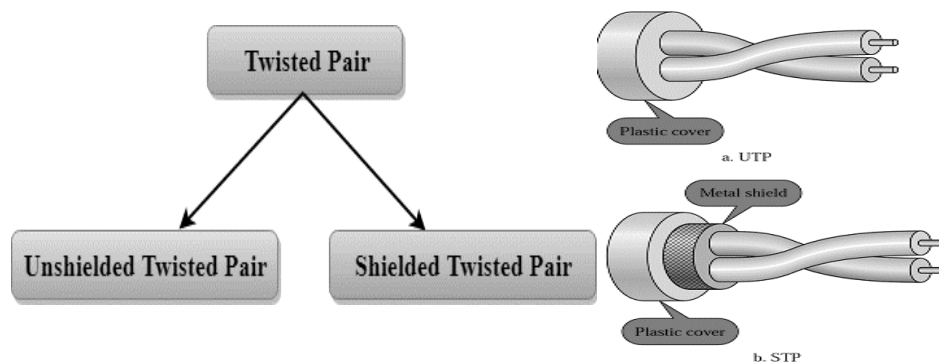


Figure 4. 5 Types of Twisted Pair

Structure of Twisted Pair

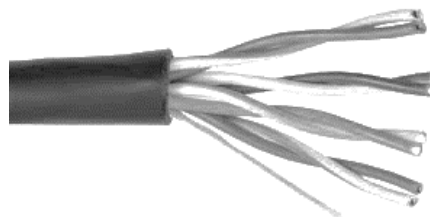


Figure 4. 6 Structure of Twisted Pair

i. Unshielded Twisted Pair

An unshielded twisted pair is widely used in telecommunication. Following are the categories of the unshielded twisted pair cable:

- **Category 1:** Category 1 is used for telephone lines that have low-speed data.
- **Category 2:** It can support upto 4Mbps.
- **Category 3:** It can support upto 16Mbps.
- **Category 4:** It can support upto 20Mbps. Therefore, it can be used for long-distance communication.
- **Category 5:** It can support upto 200Mbps.
- **Category 5e** —Used in networks running at speeds up to 1000 Mbps (1 gigabit per second [Gbps]).

- **Category 6** – Typically, Category 6 cable consists of four pairs of 24 American Wire Gauge (AWG) copper wires. Category 6 cable is currently the fastest standard for UTP.

Advantages of Unshielded Twisted Pair:

- It is cheap.
- Installation of the unshielded twisted pair is easy.
- It can be used for high-speed LAN.

Disadvantage of Unshielded Twisted Pair:

- This cable can only be used for shorter distances because of attenuation.

ii. Shielded Twisted Pair

A shielded twisted pair is a cable that contains the mesh surrounding the wire that allows the higher transmission rate.

The salient features of STP cable are:

- The Speed and throughput of the shielded Twisted pair is around 10 to 100 Mbps
- It is – Moderately expensive as compared to the Average cost per node
- Media and connector size ranges from – Medium to large
- Maximum cable length in STP is around – 100 m (short)

When you are comparing UTP and STP, you should keep the following points in mind:

- The speed of both types of cable is usually satisfactory for local-area distances.
- These are the least-expensive media for data communication. UTP is less expensive than STP.
- Because most buildings are already wired with UTP, many transmission standards are adapted to use it, to avoid costly rewiring with an alternative cable type.

Advantages of STP

- The cost of the shielded twisted pair cable is not very high and not very low.
- An installation of STP is easy.
- It has higher capacity as compared to unshielded twisted pair cable.
- It is shielded that provides the higher data transmission rate.

Disadvantages of STP

- It is more expensive as compared to UTP and coaxial cable.
 - Also it has a higher attenuation compared to the Shielded Twisted Pair.

Connectors

The most common UTP connector is RJ45 (RJ stands for registered jack). The RJ45 is a keyed connector, meaning the connector can be inserted in only one way.

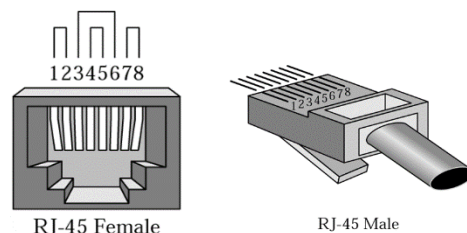


Figure 4. 7 RJ45 Connectors

b) Coaxial Cable

Coaxial cable is very commonly used transmission media, for example, TV wire is usually a coaxial cable. The name of the cable is coaxial as it contains two conductors parallel to each other. It has a higher frequency as compared to Twisted pair cable. The inner conductor of the coaxial cable is made up of copper, and the outer conductor is made up of copper mesh. The middle core is made up of non-conductive cover that separates the inner conductor from the outer conductor. The middle core is responsible for the data transferring whereas the copper mesh prevents from the EMI (Electromagnetic interference).

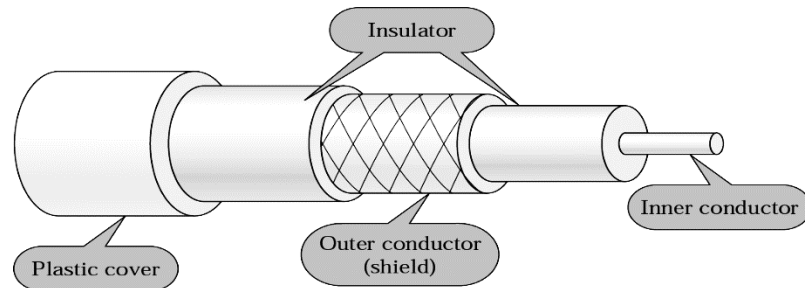


Figure 4. 8 Structure of Coaxial Cable

Types of Coaxial Cable

- Baseband transmission: It is defined as the process of transmitting a single signal at high speed.
- Broadband transmission: It is defined as the process of transmitting multiple signals simultaneously.

Advantages of Coaxial Cable

- High speed data transmission
- It has better shielding as compared to twisted pair cable.
- It provides higher bandwidth.

Disadvantages of Coaxial Cable

- It is more expensive as compared to twisted pair cable.
- If any fault occurs in the cable causes the failure in the entire network.

Connectors

- Bayonet Neill Concelman connectors are used for connection in Coaxial Cables.

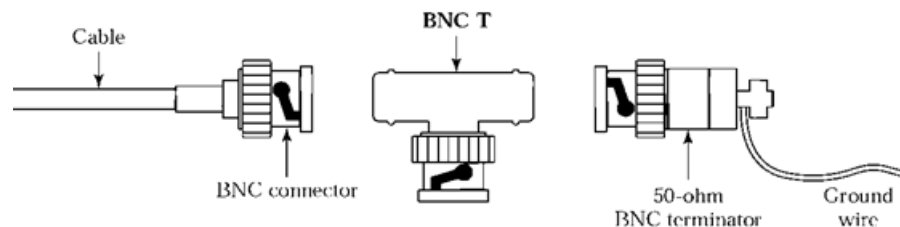


Figure 4. 9 BNC Connectors

c) Fibre Optic

Fibre optic cable is a cable that uses light signals for communication. It holds the optical fibres coated in plastic that are used to send the data by pulses of light. The plastic coating protects the optical fibres from heat, cold, electromagnetic interference from other types of wiring. Fibre optics provide faster data transmission than copper wires.

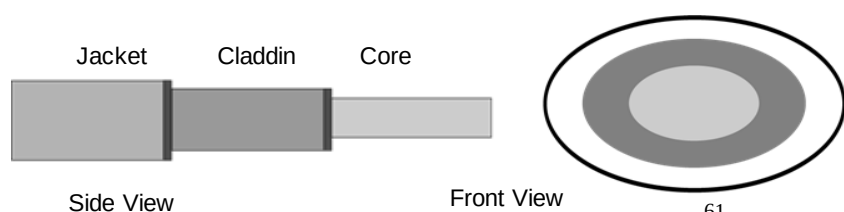


Figure 4. 10 Structure of Fiber Optics Cable

Basic Elements of Fibre Optic Cable

- **Core:** The optical fibre consists of a narrow strand of glass or plastic known as a core. A core is a light transmission area of the fibre. The more the area of the core, the more light will be transmitted into the fibre.

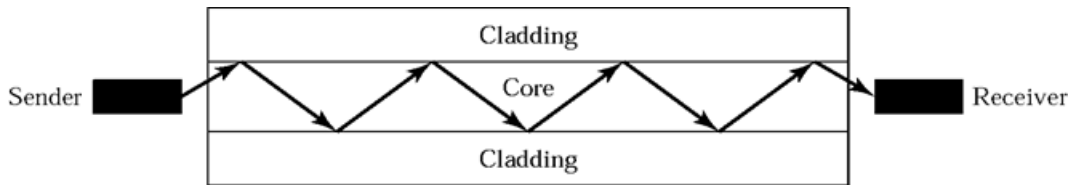


Figure 4. 11 Basic Elements of Fibre Optic Cable

- **Cladding:** The concentric layer of glass is known as cladding. The main functionality of the cladding is to provide the lower refractive index at the core interface as to cause the reflection within the core so that the light waves are transmitted through the fibre.
- **Jacket:** The protective coating consisting of plastic is known as a jacket. The main purpose of a jacket is to preserve the fibre strength, absorb shock and extra fibre protection.

How Optical Fiber Works

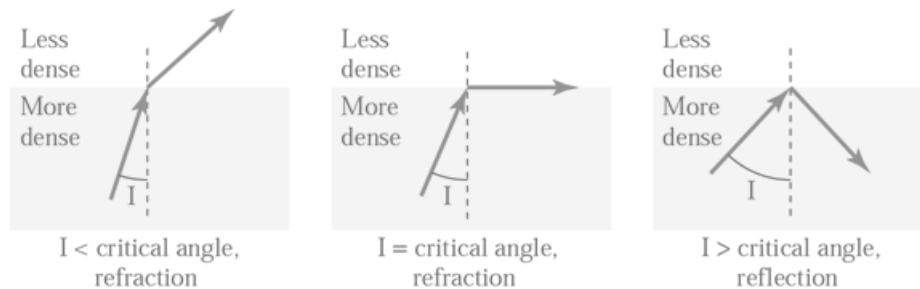


Figure 4. 12 Working of Optical Fibers

4.4 Propagation Modes

There are 2 types of propagation mode in fiber optics cable which are

- Single-mode and
 - Multi-mode
- These provide different performance with respect to both attenuation and time dispersion.
 - The single-mode fiber optic cable provides the better performance at a higher cost.

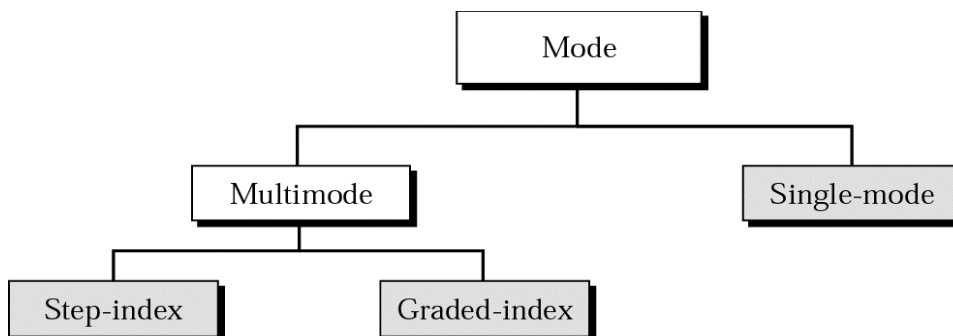


Figure 4. 13 Propagation Modes

a) Single-mode

The diameter of the core is fairly small relative to the cladding. Typically, the cladding is ten times thicker than the core. When fiber core radius is reduced, fewer angles will reflect By

reducing the radius of the core to the order of a wavelength, only a single angle or mode can pass – the axial ray. Single mode propagation exists only above a certain specific wavelength called the cutoff wavelength.



Figure 4. 14 Single Mode

b) Multi-mode

i. Multi-mode Step Index

Refers to the variety of angles that will reflect. Multiple propagation path exists, signal elements spread out in time and hence the data rate.

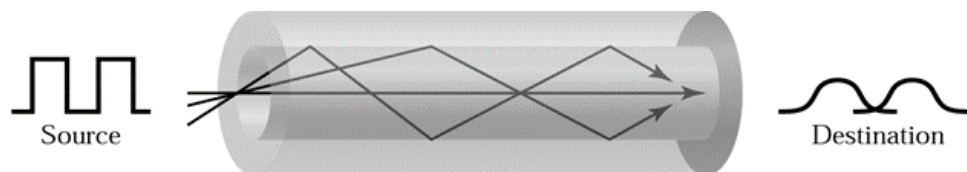


Figure 4. 15 Multimode, Step Index

ii. Multi-mode Graded Index

By varying the refractive index of the core, rays may be focused more efficiently than multimode.

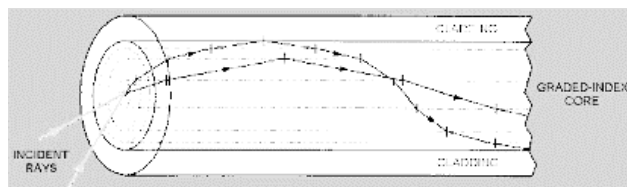


Figure 4. 16 Multimode Graded Index

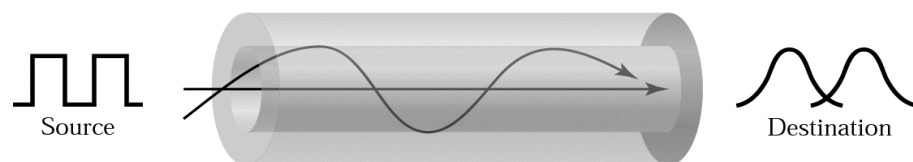


Figure 4. 17 Multimode Graded Index

Advantages of fibre optic cable over copper

- **Greater Bandwidth:** The fibre optic cable provides more bandwidth as compared copper. Therefore, the fibre optic carries more data as compared to copper cable.
- **Faster speed:** Fibre optic cable carries the data in the form of light. This allows the fibre optic cable to carry the signals at a higher speed.
- **Longer distances:** The fibre optic cable carries the data at a longer distance as compared to copper cable.
- **Better reliability:** The fibre optic cable is more reliable than the copper cable as it is immune to any temperature changes while it can cause obstruct in the connectivity of copper cable.
- **Thinner and Sturdier:** Fibre optic cable is thinner and lighter in weight so it can withstand more pull pressure than copper cable.

2. Un-Guided Transmission

An unguided transmission transmits the electromagnetic waves without using any physical medium. Therefore it is also known as **wireless transmission**. In unguided media, air is the media through which the electromagnetic energy can flow easily. Unguided transmission is broadly classified into three categories:

- a) Radio waves
- b) Microwaves
- c) Infrared

a) Radio waves

Radio waves are the electromagnetic waves that are transmitted in all the directions of free space. Radio waves are omnidirectional, i.e., the signals are propagated in all the directions. The range in frequencies of radio waves is from 3Khz to 1 khz. In the case of radio waves, the sending and receiving antenna are not aligned, i.e., the wave sent by the sending antenna can be received by any receiving antenna.



An example of the radio wave is **FM radio**.

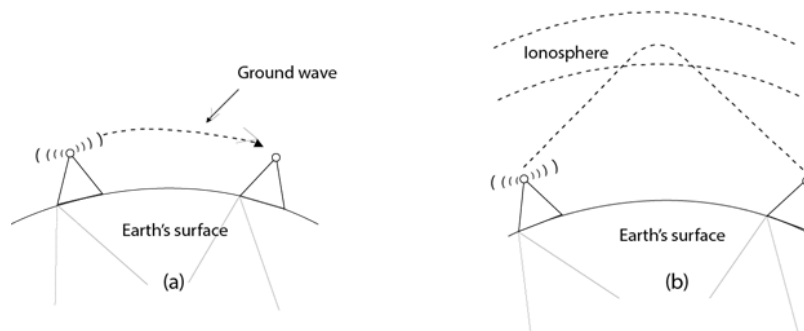


Figure 4.18 Ground Wave and Skywave

Applications of Radio Waves

- A Radio wave is useful for multicasting when there is one sender and many receivers.
- An FM radio, television, cordless phones are examples of a radio wave.

Advantages of Radio Transmission

- Radio transmission is mainly used for wide area networks and mobile cellular phones.
- Radio waves cover a large area, and they can penetrate the walls.
- Radio transmission provides a higher transmission rate.

Modes of Propagation

- In Radio communication systems, we use wireless electromagnetic waves as the channel. The antennas of different specifications can be used for these purposes.
- The mode of propagation of electromagnetic waves in the atmosphere and in free space may be divided into the following three categories:
 - i. The line of sight (LOS) propagation
 - ii. Ground wave propagation
 - iii. Skywave propagation

i. Line of Sight (LOS) Propagation

In the line-of-sight communication, as the name implies, the wave travels a minimum distance of sight. Which means it travels to the distance up to which a naked eye can see. Then we need to employ an amplifier cum transmitter here to amplify the signal and transmit again. The line-

of-sight propagation will not be smooth if there occurs any obstacle in its transmission path. As the signal can travel only to lesser distances in this mode, this transmission is used for infrared or microwave transmissions.

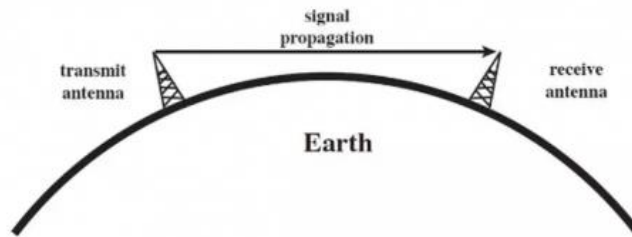


Figure 4. 19 Line of Sight Propagation (above 30 MHz)

ii. Ground Wave Propagation

Ground wave propagation of the wave follows the contour of the earth. Such a wave is called a direct wave. The wave sometimes bends due to the Earth's magnetic field and gets reflected the receiver. Such a wave can be termed as a reflected wave. The following figure depicts ground wave propagation. The wave then propagates through the Earth's atmosphere is known as a ground wave. The direct wave and reflected wave together contribute the signal at the receiver station. When the wave finally reaches the receiver, the lags are cancelled out. In addition, the signal is filtered to avoid distortion and amplified for clear output.

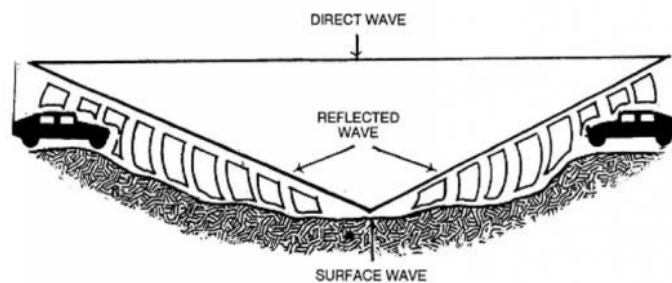


Figure 4. 20 Components of Ground Wave

iii. SkyWave Propagation

Skywave propagation is preferred when the wave has to travel a longer distance. Here the wave is projected onto the sky and it is again reflected back to the earth. The waves, which are transmitted from the transmitter antenna, are reflected from the ionosphere. It consists of several layers of charged particles ranging in altitude from 30-250 miles above the surface of the earth. Such travel of the wave from the transmitter to the ionosphere and from there to the receiver on Earth is known as Sky Wave Propagation. The ionosphere is the ionized layer around the Earth's atmosphere, which is suitable for skywave propagation.

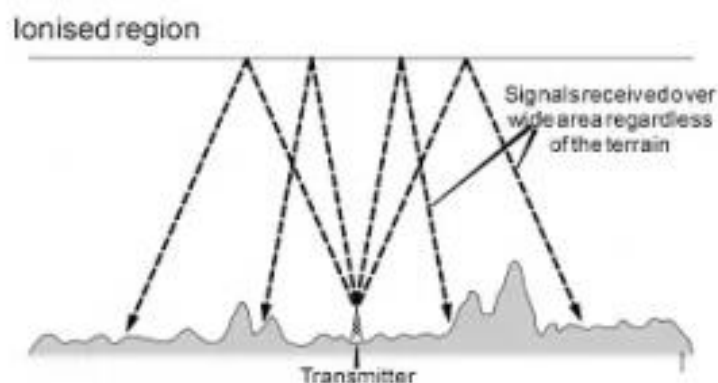


Figure 4. 21 Skywave Propagation

b) Microwaves

- Microwaves are of two types:
 - Terrestrial microwave
 - Satellite microwave communication

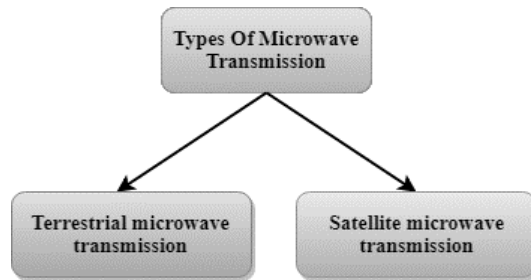


Figure 4. 22 Types of Microwave Transmission

i. Terrestrial Microwave Transmission

Terrestrial Microwave transmission is a technology that transmits the focused beam of a radio signal from one ground-based microwave transmission antenna to another. Microwaves are the electromagnetic waves having the frequency in the range from 1GHz to 1000 GHz. Microwaves are unidirectional as the sending and receiving antenna is to be aligned, i.e., the waves sent by the sending antenna are narrowly focussed. In this case, antennas are mounted on the towers to send a beam to another antenna which is km away. It works on the line of sight transmission, i.e., the antennas mounted on the towers are the direct sight of each other.

Characteristics of Terrestrial Microwave:

- Frequency range: The frequency range of terrestrial microwave is from 4-6 GHz to 21-23 GHz.
- Bandwidth: It supports the bandwidth from 1 to 10 Mbps.
- Short distance: It is inexpensive for short distance.
- Long distance: It is expensive as it requires a higher tower for a longer distance.
- Attenuation: Attenuation means loss of signal. It is affected by environmental conditions and antenna size.

Advantages of Terrestrial Microwave:

- Microwave transmission is cheaper than using cables.
- It is free from land acquisition as it does not require any land for the installation of cables.
- Microwave transmission provides an easy communication in terrains as the installation of cable in terrain is quite a difficult task.
- Communication over oceans can be achieved by using microwave transmission.

Disadvantages of Terrestrial Microwave:

- Eavesdropping: An eavesdropping creates insecure communication. Any malicious user can catch the signal in the air by using its own antenna.

4.5 Out of phase signal

A signal can be moved out of phase by using microwave transmission. A microwave transmission is susceptible to weather condition. This means that any environmental change such as rain, wind can distort the signal. Allocation of bandwidth is limited in the case of microwave transmission.

ii. Satellite Microwave Communication

- A satellite is a physical object that revolves around the earth at a known height.

- Satellite communication is more reliable nowadays as it offers more flexibility than cable and fibre optic systems.
- We can communicate with any point on the globe by using satellite communication.

How Does Satellite work?

- The satellite accepts the signal that is transmitted from the earth station, and it amplifies the signal. The amplified signal is retransmitted to another earth station.

Advantages of Satellite Microwave Communication:

- The coverage area of a satellite microwave is more than the terrestrial microwave.
- The transmission cost of the satellite is independent of the distance from the centre of the coverage area.
- Satellite communication is used in mobile and wireless communication applications.
- It is easy to install.
- It is used in a wide variety of applications such as weather forecasting, radio/TV signal broadcasting, mobile communication, etc.

Disadvantages of Satellite Microwave Communication:

- Satellite designing and development requires more time and higher cost.
- The Satellite needs to be monitored and controlled on regular periods so that it remains in orbit.
- The life of the satellite is about 12-15 years. Due to this reason, another launch of the satellite has to be planned before it becomes non-functional.

c) Infrared

An infrared transmission is a wireless technology used for communication over short ranges. The frequency of the infrared is in the range from 300 GHz to 400 THz. It is used for short-range communication such as data transfer between two cell phones, TV remote operation, data transfer between a computer and cell phone resides in the same closed area.

Characteristics of Infrared

- It supports high bandwidth, and hence the data rate will be very high.
- Infrared waves cannot penetrate the walls. Therefore, the infrared communication in one room cannot be interrupted by the nearby rooms.
- An infrared communication provides better security with minimum interference.
- Infrared communication is unreliable outside the building because the sun rays will interfere with the infrared waves.

4.6 Introduction to Networking Devices

LANs do not normally operate in isolation but they are connected to one another or to the Internet. To connect LANs, connecting devices are needed and various connecting devices are such as bridge, switch, router, hub, repeater. These connecting devices are Networking Devices

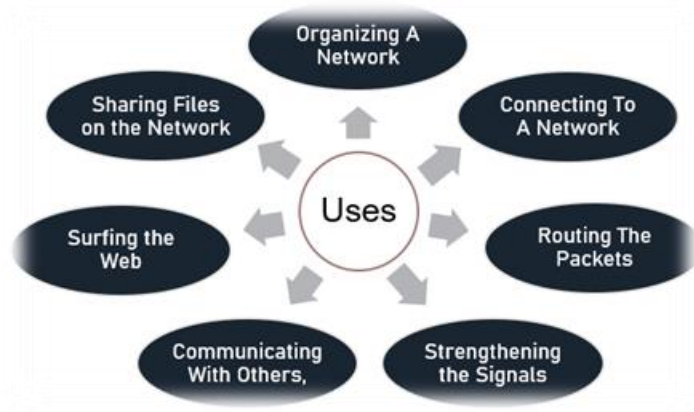


Figure 4. 23 Uses of Networking Devices

Connecting Devices

Connecting devices into five different categories based on the layer in which they operate in a network.

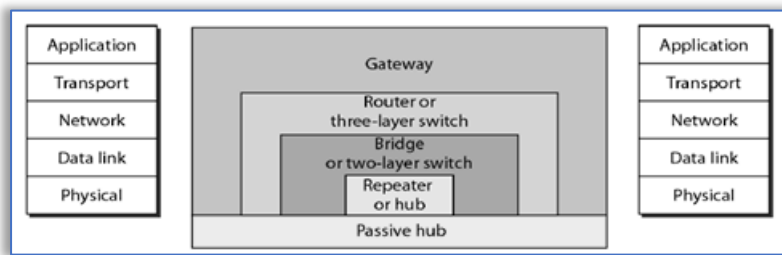


Figure 4. 24 Five Categories of Networking Devices

Types of Network Devices

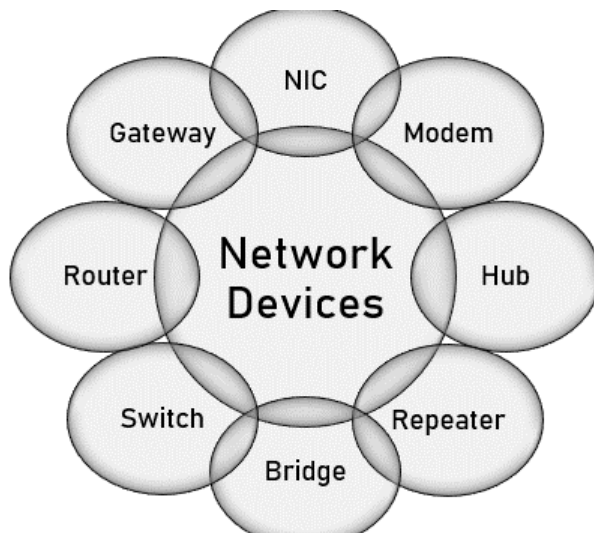


Figure 4. 25 Types of Networking Devices

a) Network Interface Card (NIC)

A network interface controller is a computer hardware component that connects a computer to a computer network. Early network interface controllers were commonly implemented on expansion cards that plugged into a computer bus.

A Network Interface Card (NIC) is circuit board or a card that allows computers to communicate over a network via cables or wirelessly.

It is also called as LAN adaptor, network adaptor or network card. It enables clients, servers, printers and other devices to transmit and receive data over the network. It operates on physical and data link layer of OSI model. Every network adaptor is assigned a unique 48-bit Media Access Control (MAC) address, which is stored in ROM to identify themselves in a network or a LAN. The available maximum data transfer rate is 10, 100 and 1000 MBPS. Typically network adaptor has RJ45 or BNC or both sockets for connecting and a LED to show up it is active and transmitting the data. It connects to a network via cables like CAT5, Co-axial, fibre-optics etc. and wirelessly by a small antenna.

b) *Modem*

Modem stands for "modulator-demodulator" – is a hardware device that converts data from a digital format, intended for communication directly between devices with specialized wiring, into one suitable for a transmission medium such as telephone lines or radio. A modem modulates one or more carrier wave signals to encode digital information for transmission and demodulates signals to decode the transmitted information. The goal is to produce a signal that can be transmitted easily and decoded reliably to reproduce the original digital data. It is a device that converts digital signal to analog signal as a modulator and analog signal to digital signal as a demodulator. It enables computers to communicate over telephone lines. Speed of modem is measured in bits per second and varies depending upon the type of modem. Higher the speed, the faster you can send and receive data over the network. It is used to connect computer to the internet.

Working of Modem

- Consider a communication between two computers A and B.
- Computer A transmits the digital signals to its modem in the form of binary 0's and 1's.
- Modem of computer A converts these digital signals into analog signals and sends over the telephone line. This process is called as modulation.
- While at the other end, modem of computer B receives the analog signals and converts back into digital signals. This process is called as demodulation.
- Converted digital signals by the modem are sent to the computer B for processing.
- In similar way computer B can communicate with computer A.

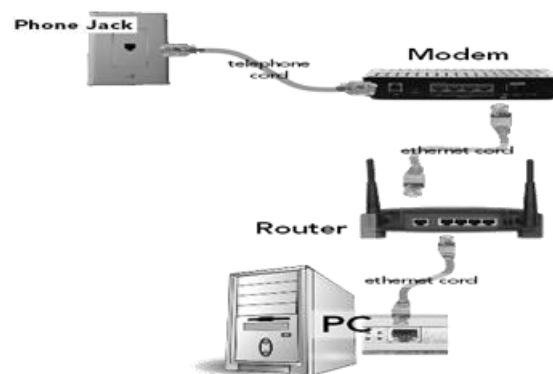


Figure 4. 26 Working of Modem

c) *Hub*

An Ethernet hub, active hub, network hub, repeater hub, multiport repeater, or simply hub is a network hardware device for connecting multiple Ethernet devices together and making them act as a single network segment. Hubs are devices commonly used to connect segments of a LAN. A hub has many ports in it. A computer which intends to be connected to the network is plugged in to one of these ports. When a data frame arrives at a port, it is broadcast to every other port, without considering whether it is destined for a particular destination or not.

Types of Hubs

1. Passive Hub

This type of does not amplify or boost the signal. It does not manipulate or view the traffic that crosses it. The passive hub does not require electrical power to work.

2. Active Hub

It amplifies the incoming signal before passing it to the other ports. It requires AC power to do the task.

3. Intelligent Hub

They are also called as smart hubs. Function as an active hub and include diagnostic capabilities. Intelligent hubs include microprocessor chip and are very useful in troubleshooting conditions of the network.

d) Repeaters

A repeater is a device that operates only at the Physical Layer. It can be used to increase the length of the network by eliminating the effect of attenuation on the signal. It connects two segments of the same network, overcoming the distance limitations of the transmission media. A repeater forwards every frame; it has no filtering capability. It is a regenerator, not an amplifier. It can connect segments that have the same access method. (CSMA/CD, Token Passing, Polling, etc.)

Functions of a Repeater

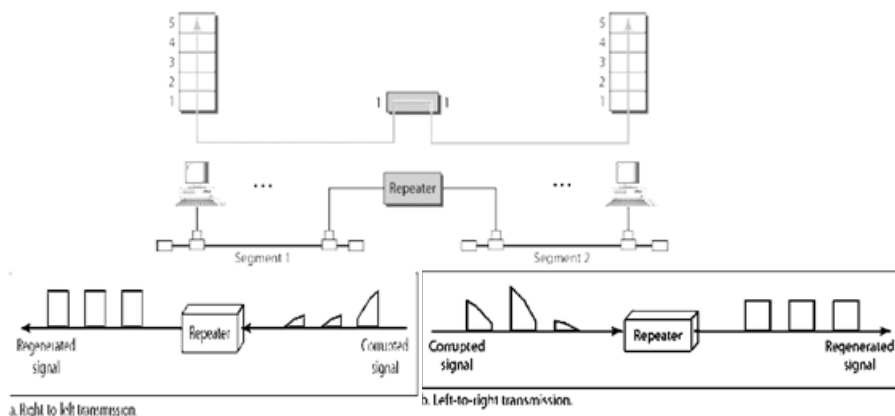


Figure 4. 27 Functions of Repeaters

e) Bridges

Bridges operate in both the Physical and the Data Link Layer. As a Physical Layer device, it regenerates the signal it receives. As a Data Link layer device, the bridge can check the Physical/MAC addresses (source and destination) contained in the frame. A bridge has a table used in filtering decisions. It can check the destination address of a frame and decide if the frame should be forwarded or dropped. If the frame is to be forwarded, the decision must specify the port. A bridge has a table that maps addresses to ports. Limits or filters traffic keeping local traffic local yet allow connectivity to other parts (segments).

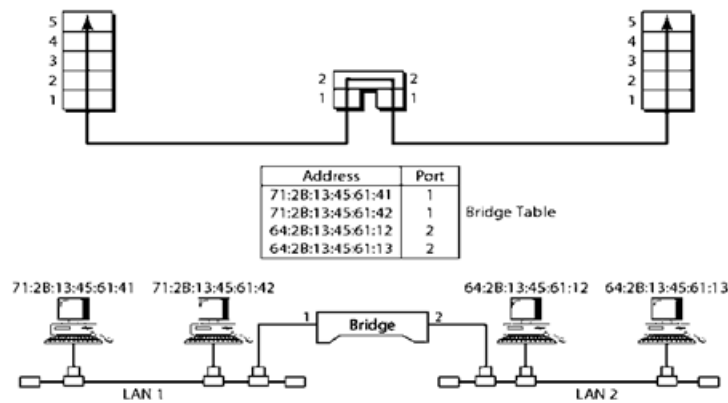


Figure 4. 28 A Bridge Connected to Two LANs

A bridge does not change the physical MAC address

Characteristics of Bridges

- **Routing Tables**

Contains one entry per station of network to which bridge is connected. It is used to determine the network of destination station of a received packet.

- **Filtering**

Is used by bridge to allow only those packets which are destined to the remote network

Packets are filtered with respect to their destination and multicast addresses.

- **Forwarding**

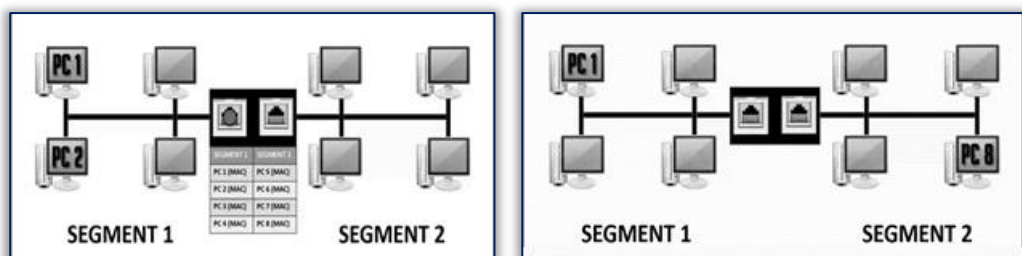
The process of passing a packet from one network to another.

- **Learning Algorithm**

The process by which the bridge learns how to reach stations on the internetwork

How Bridges Work

- Bridges work at the Media Access Control Sub-layer of the OSI model.
- Routing table is built to record the segment number of address.
- If destination address is in the same segment as the source address, stop the transmit.
- Otherwise forward to the other segment.



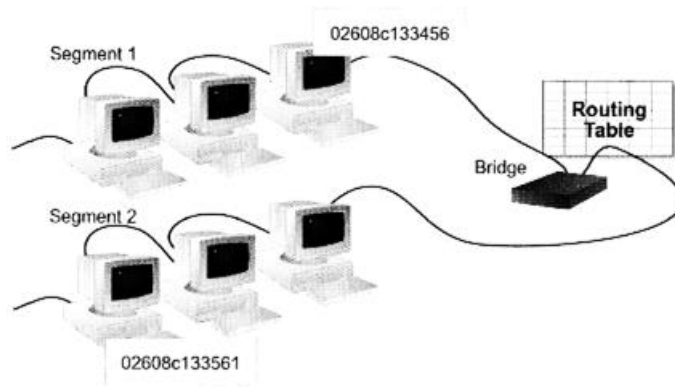
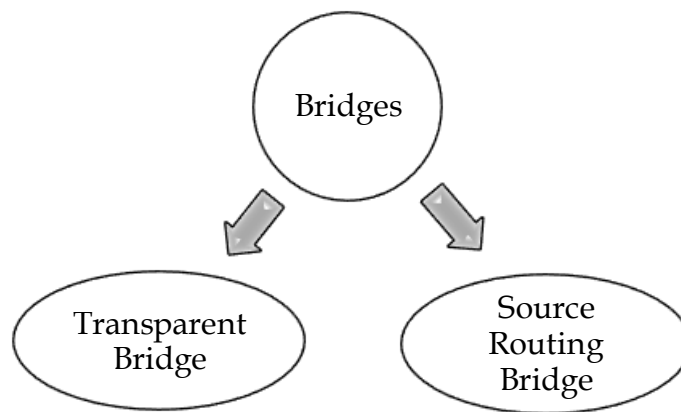


Figure 4. 29 Working of Bridge with Multiple Segments

Types of Bridges



Transparent Bridge

- Also called Learning Bridges.
- Build a table of MAC addresses as frames arrive.
- Ethernet networks use transparent bridge.

Duties of transparent bridge are:

- Filtering
- Framing
- Forwarding
- Blocking

Source Routing Bridge

- Used in token Ring networks.
- Each station should determine the route to the destination when it wants to send a frame and therefore includes the route information in the header of frame.
- Address of these bridges are included in the frame.
- Frame contains the bridge addresses as well apart from the source and destination address.

Advantages of Bridges

- Extended physical network
- Reduces network traffic

- Creates separate collision domains
- Reduces collisions
- Connect different architecture

Disadvantages of Bridges

- Slower than repeaters due to filtering
- Do not filter broadcasts
- More expensive than filters

f) Switch

Switches allow different devices on a network to communicate. A Network Switch is a constituent of computer network that connects two network slices and/or two network devices (switches or routers) together. It can be termed as a network bridge with multiple ports which helps to process and route packets at data link layer of the OSI reference model.

Working

Switch use two different methods for switching the packets

- i. Cut-Through Method
- ii. Store and Forward Method

i. Cut-Through Method

In this method switch examines the header of the packet and decides, where to pass the packet before it receives the whole packet. Increases the chances of errors without verifying the data integrity.

ii. Store and Forward Method

In this method switch reads the entire packet in its memory and checks for error before transmitting the packet. This method is slower and time consuming but error free.

g) Router

A router is a network layer hardware device that transmits data from one LAN to another if both networks support the same set of protocols.

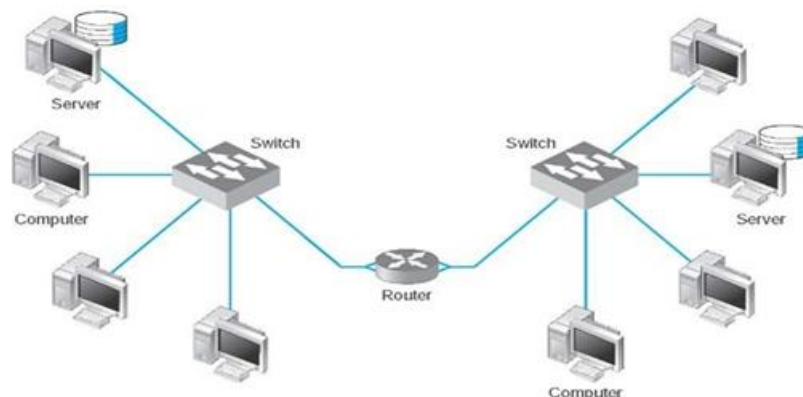


Figure 4. 30 A Router Connecting two LAN's

So, a router is typically connected to at least two LANs and the internet service provider (ISP). It receives its data in the form of packets, which are data frames with their destination address added. A router reads its routing table to decide the best available route the packet can take to reach its destination quickly and accurately.

The routing table may be of these two types:

- **Static Routing Table**

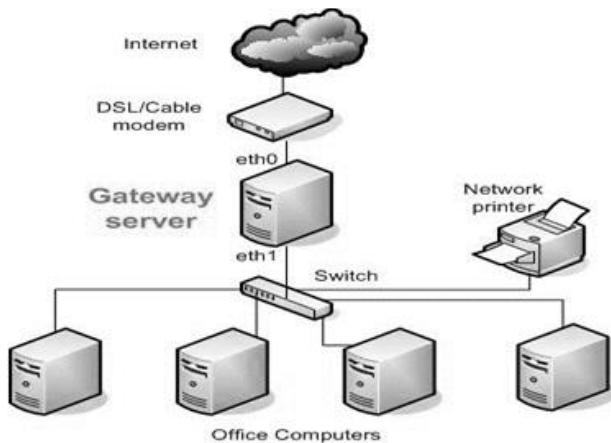
In a static routing table the routes are fed manually. So, it is suitable only for very small networks that have maximum two to three routers.

- **Dynamic Routing Table**

In a dynamic routing table, the router communicates with other routers through protocols to determine which routes are free. This is suited for larger networks where manual feeding may not be feasible due to large number of routers.

h) Gateway

Gateway is a network point that act as entry point to other network and translates one data format to another.



Functions of the Gateway

- Protocol translation – It translates protocol format into required protocol format of the network, such as X.25 to TCP/IP.
- Network address translation – It translates your public IP address to the private IP addresses on your network.
- DHCP service – It automatically assigns IP address to a computer from a defined range of addresses for a given network.
- Monitoring and regulating each packet entering and leaving the network.



Give a brief description of the application and limitations of the following types of transmission media:

- Two-wire open lines
- Twisted pair lines
- Coaxial cable
- Optical fiber
- Microwaves

Summary

- There are several kinds of transmission media. These media technologies starting from copper wire to wireless and fiber optic has grown-up so rapidly and replacing other very quickly in this information age.
- Transmission media can be broadly classified into two types: Guided and Unguided transmission media.

- Twisted pair, coaxial cable, and optical fiber fall into the category of guided or bounded transmission media.
- Twisted pair is a pair of copper wires twisted together and wrapped with a plastic coating. It is mainly of two types: shielded twisted pair (STP) and Unshielded twisted pair (UTP).
- Shielded twisted pair (STP) differs from UTP in that a metallic shield or screen surrounds the pairs, which may or may not be twisted.
- Coaxial Cable is a very robust shielded copper wire two-conductor cable in which a solid center conductor runs concentrically (coaxial) inside a solid outer circular conductor.
- Optical fiber carries the transmitted information in the form of a fluctuating beam of light in a glass fiber rather than as an electrical signal on a wire. It can be of two types: monomode and multimode fiber.

Keywords

Bandwidth: Refers to the range of frequencies assigned to a channel.

Bounded Media: Refers to the wired transmission systems that employ physical media, which are tangible.

Coaxial Cable: It is a very robust shielded copper wire two-conductor cable in which a solid center conductor runs concentrically (coaxial) inside a solid outer circular conductor.

Frequency Spectrum: Refers to the range of frequencies being supported by a particular transmission medium.

Gauge: Gauge is a measure of the thickness of the conductor.

Graded Index Multimode Fiber: In the case of a graded index multimode fiber, the index of refraction across the core is gradually changed from a maximum at the center to a minimum near the edges, hence the name graded index.

Monomode/Singlemode fiber: This has a thinner inner core. In this case, the core diameter of about 9 μm is much closer in size to the wavelength of light being propagated, about 1.3 μm . This limits the light transmission to a single ray or mode of light to propagate down the core of the fiber.

Multimode Fiber: The core diameter is relatively large compared to a wavelength of light.

Optical Fiber: Optical fiber carries the transmitted information in the form of a fluctuating beam of light in a glass fiber rather than as an electrical signal on a wire.

Propagation Delay: Refers to the length of time required for a signal to travel from transmitter to receiver across a transmission system.

Shielded Copper or STP: Shielded twisted pair (STP) differs from UTP in that a metallic shield or screen surrounds the pairs, which may or may not be twisted.

Step Index Multimode Fiber: Different rays travel different distances, and take different amounts of time to transit the length of a fiber.

Twisted Pair: A twisted pair is a pair of copper wires twisted together and wrapped with a plastic coating.

Unbounded Media: Refers to wireless transmission systems do not make use of a physical conductor, or guide, to bind the signal.

Unshielded Twisted Pair (UTP): A UTP cable contains from 2 to 4200 twisted pairs. The advantages of UTP are the flexibility, low cost media, and can be used for either voice or data communications.

Self Assessment

Fill in the blanks:

1. can be broadly categorized into guided and unguided media.
2. The actual range of frequencies supporting a given communication is known as a
3. In general, the higher the, the more will be the data transmission rate or throughput.
4. refers to the length of time required for a signal to travel from transmitter to receiver across a transmission system.
5. Bandwidth may be defined as the range ofassigned to a channel.

State whether the following statements are true or false:

6. Twisted pair (both unshielded and shielded), coaxial and fiber optic cable systems fall into ;guided transmission media category.
7. The twisting decreases the electrical noise immunity, and reduces the error rate of the data transmission.
8. A UTP cable contains from 2 to 4200 twisted pairs.
9. Coaxial cable is inherently an insecure transmission medium.
10. Local Area Networks can operate over coaxial cable to the 10BASE5, 10BASE2 and 10BASET specifications

Review Questions

1. What are the different transmission mediums over which data communication devices can provide service?
2. What are the major limitations of twisted pair wire?
3. Describe how satellite communication is different from radio broadcast?
4. State with the help of a diagram the different components of typical fiber optic link. Mention the various components of signal loss.
5. What is reflection? What happens to a beam of light as it travels to a less dense medium? What happens if it travels to a denser medium?
6. What advantages do coaxial cables offer over twisted pair cables?
7. Compare fiber optic cable with UTP cable when used as transmission media in LANs.
8. What is the purpose of cladding in an optical fiber? Discuss its density with respect to the core.
9. What is skin effect and how does it affect the performance of TP cables?
How does coaxial cable reduce the problem of skin effect and becomes an appropriate media for higher frequency data transmission?
10. Which type of transmission media does find extensive deployment for digital transmission and why?

Answers: Self Assessment

- | | |
|-----------------------|----------------------|
| 1. Transmission media | 2. Pass band |
| 3. Bandwidth | 4. Propagation delay |
| 5. Frequencies | 6. True |
| 7. False | 8. True |
| 9. False | 10. True |

Further Readings



Andrew S. Tanenbaum, *Computer Networks*, Prentice Hall

Behrouz A. Forouzan and Sophia Chung Fegan, *Data Communications and Networking*, McGraw-Hill Companies

Burton, Bill, *Remote Access for Cisco Networks*, McGraw-Hill Osborne Media McGraw-Hill Osborne Media

Rajneesh Agrawal and Bharat Bhushan Tiwari, *Computer Networks and Internet*, Vikas Publication



<https://www.geeksforgeeks.org/basics-computer-networking/>

Unit 05: Data Link Layer - Error Detection and Correction Methods

CONTENTS

Objectives

Introduction

5.1 Understanding the Data Link Layer

5.2 Error Detection and Correction

5.2.1 Error Correction

Summary

Keywords

Self Assessment

Review Questions

Answers: Self Assessment

Further Readings

Objectives

After this lecture, you would be able to

- understand the various services provided by the Data Link Layer
- learn the Logical Link Layer Services like the framing, error control and flow control.
- understand the concept of Block Coding
- learn about the various types of errors and their possible causes
- learn about the various error detection and correction codes.
- understand the use of hamming code for error detection and correction
- learn the technique by practicing a few numerical problems

Introduction

Data link layer is the second layer after physical layer in the OSI reference model. It describes the techniques to access a shared communication channel and reliable transmission of data frame in computer communication environment. It receives a raw stream of bits for the physical layer at sender machine. The raw stream of data is created using different technologies like cable, DSL, wireless, optical fiber, etc. The data link layer transforms data free of undetected transmission errors to the network layer. Data link layer accomplishes this task by using acknowledgment frames and error detection algorithms. In other words, the task of the data link layer is to transmit the bits to the destination machine. The data link layer of the destination machine, then, hand over thus received data to the network layer for processing.

5.1 Understanding the Data Link Layer

The data link layer is the protocol layer in a program that handles the moving of data into and out of a physical link in a network. The data link layer is Layer 2 in the Open Systems Interconnection (OSI) architecture model for a set of telecommunication protocols. It ensures that the error free data is transferred between the adjacent nodes in the network.

Functions of Data Link Layer

Data Link Layer provides two main functionalities:

- It provides a reliable data transfer service between two peer network layers

- Providing a well-defined service interface to the network layer.
- The main responsibility of the Data Link Layer is to deal with the transmission errors.
- Flow Control mechanism which regulates the flow of frames such that data congestion should not occur at slow receivers due to fast senders.

The basic functions of the Data Link Layer are:

- a. Framing
- b. Physical Addressing
- c. Synchronization
- d. Error Control
- e. Flow Control and
- f. Multiple-Access

Figure 5.1 illustrates the functions of the Data Link Layer.

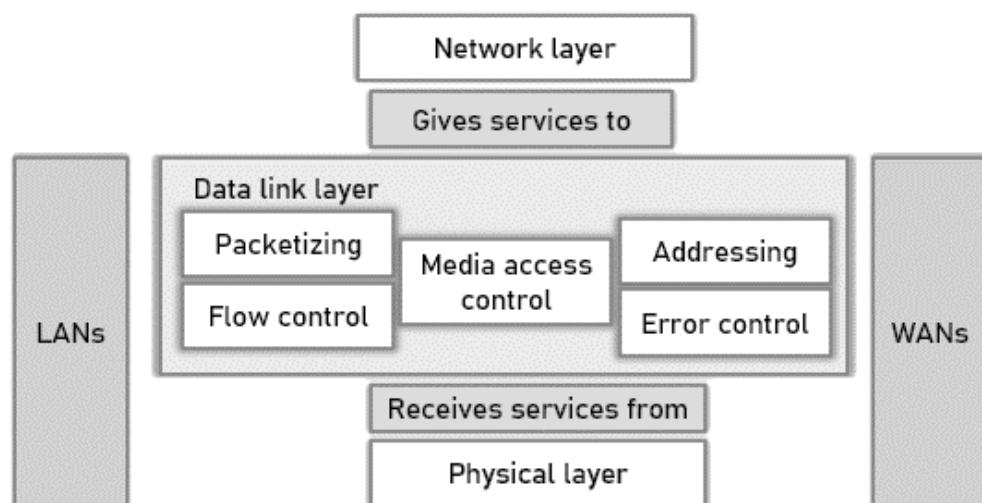


Figure 5. 1 Functions of the Data Link Layer

a. Framing

Frames are the units of digital transmission particularly in computer networks and telecommunications. Framing is a function of the data link layer. It provides a way for a sender to transmit a set of bits that are meaningful to the receiver.

Data-link layer takes packets from Network Layer and encapsulates them into Frames. Then, it sends each frame bit-by-bit on the hardware. At receiver' end, data link layer picks up signals from hardware and assembles them into frames.

Ethernet, token ring, frame relay, and other data link layer technologies have their own frame structures. Frames have headers that contain information such as error-checking codes.

b. Physical Addressing

Data-link layer provides layer-2 hardware addressing mechanism. Hardware address is assumed to be unique on the link. It is encoded into hardware at the time of manufacturing.

c. Synchronization

For a data frame be sent on the link and the data transfer to take place, both the machines need to be synchronized. This essential task is also performed by the Data Link Layer

d. Error Control

During data transmission, some bits may get flipped. This creates problems in the transmission of signals. On detecting these errors, it attempts to recover the actual data bits. It also lays down a mechanism to report errors to the sender.

e. Flow Control

To understand flow control let us take an example. Remember in your childhood when your mother used to feed you food. She used to give you bites and waited for you to finish eating it and only then she would give you the next bite of food. If she would have kept on feeding you without waiting for you to finish which you confirmed by nodding your head. This would lead you to spill everything on the floor. Similarly, in case of networks, stations on same link may have different speed or capacity. Data-link layer ensures flow control that enables both machine to exchange data on same speed.

f. Multiple-Access

When working on multiple systems, let us consider a scenario. When a host on the shared link tries to transfer the data, there is a high probability of collision. Data-link layer provides mechanism such as CSMA/CD to equip the capability of accessing shared media among multiple Systems.

Therefore, the two main functions of the data link layer are:

- **Logical Link Control (LLC):** It deals with the design and procedures for communication b/w nodes: node-to-node communication.
- **Media Access Control (MAC):** It explains how to share the link.

The functions can be seen in Figure 5.2.

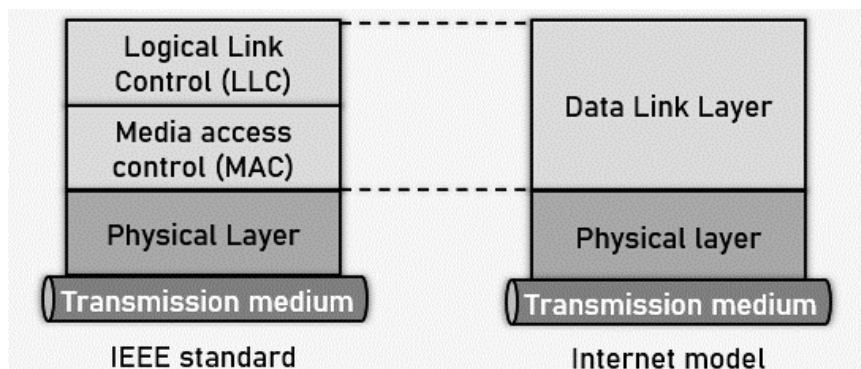


Figure 5. 2 Two main functions of Data Link Layer

Logical Link Control

- Logical link control functions includes various functions like framing, flow control, error control, and software implemented protocols for the smooth and reliable transmission of frames between the different nodes.
- To implement logical link control, we need protocols.

As you already know, Protocols are a set of rules which need to be implemented in software and are run by the two nodes involved in data exchange at the data link layer. The diagram clearly shows the various components of the Logical Link Control Layer. The main components are 802.3(CSMA/CD), 802.4(Token Bus), 802.5(Token Ring), 802.6(DQBA) and the 802.11(Wireless) standard to name a few.

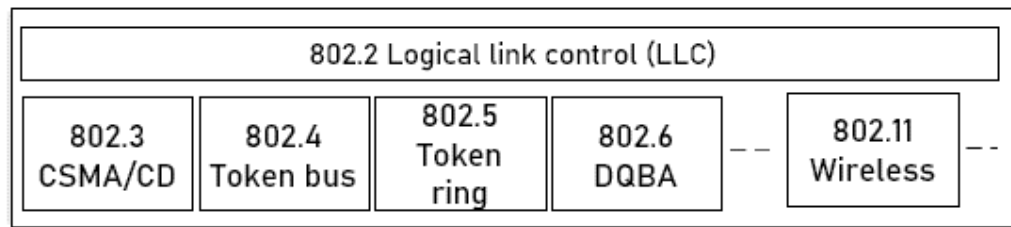


Figure 5. 3 Main components are 802.3 Logical Link Control

Logical Link Layer Services

- a) Framing
- b) Error Control
- c) Flow Control

a) Framing

The data link layer needs to pack bits into frames, so that each frame is distinguishable from another.

Our postal system practices a type of framing. The simple act of inserting a letter into an envelope separates one piece of information from another; the envelope serves as the delimiter. This can be seen in Figure 5.3.

The frame contains

- Frame header
- Payload field for holding packet
- Frame trailer

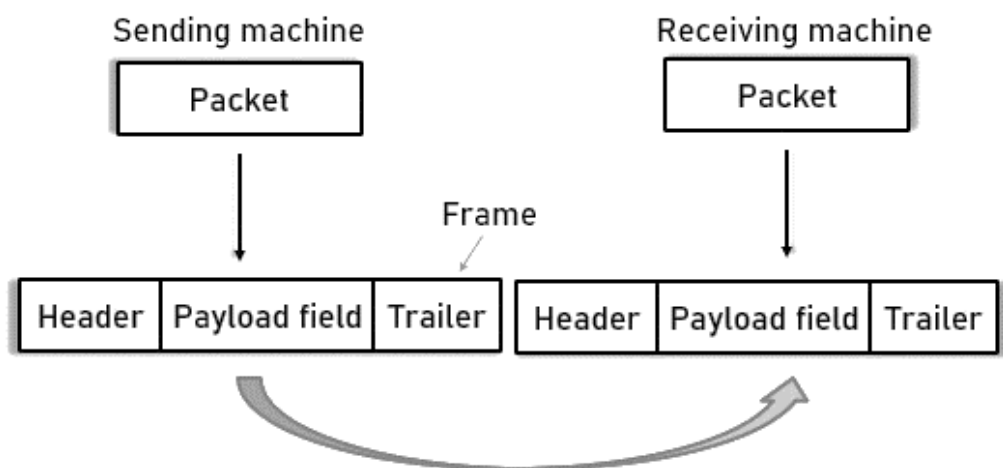


Figure 5. 4 Components of Frame and its Working

During transmission, the bit streams need to be broken up into frames. This is a difficult task. This could be achieved in different ways. One way to achieve this framing is to insert time gaps between frames, much like the spaces between words in ordinary text. However, networks rarely make any guarantees about timing, so it is possible these gaps might be squeezed out or other gaps might be inserted during transmission.

Now it is time for us to discuss how Framing can be done. So, There are four methods:

- i Character count
- ii Flag bytes with byte stuffing
- iii Starting and ending flags, with bit stuffing
- iv Physical layer coding violations

i. Character count

The first framing method uses a field in the header to specify the number of characters in the frame. When the data link layer at the destination sees the character count, it knows how many characters follow and hence where the end of the frame is.

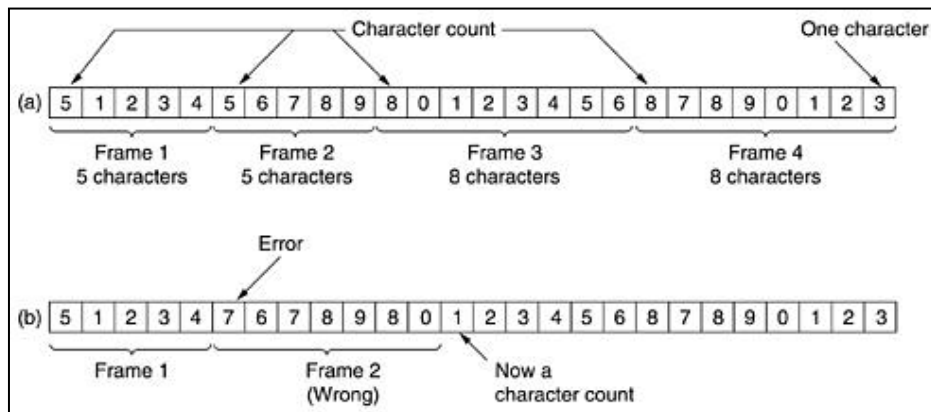


Figure 5.5 Flag bytes with Byte Stuffing

This type of approach is also known as character-oriented approach, data to be carried are 8-bit characters. The header, which normally carries the source and destination addresses and other control information. Trailer carries error detection or error correction redundant bits, are also multiples of 8 bits. This is evident from Figure 5.6.

ii. Flag bytes with byte stuffing

To separate one frame from the next, an 8-bit (1-byte) flag is added at the beginning and the end of a frame. The flag, composed of protocol-dependent special characters, signals the start or end of a frame. (as shown in Figure 5.5)

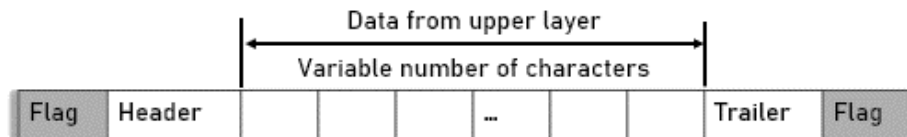


Figure 5.6 Character-oriented Approach

Byte Stuffing and Unstuffing

The Byte Stuffing and Unstuffing can be understood with the help of Figure 5.7

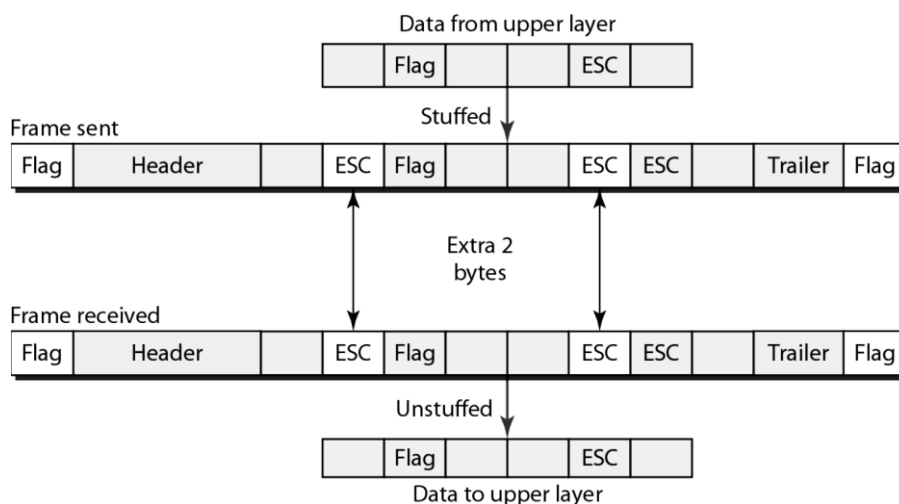


Figure 5.7(a) Byte Stuffing and Unstuffing

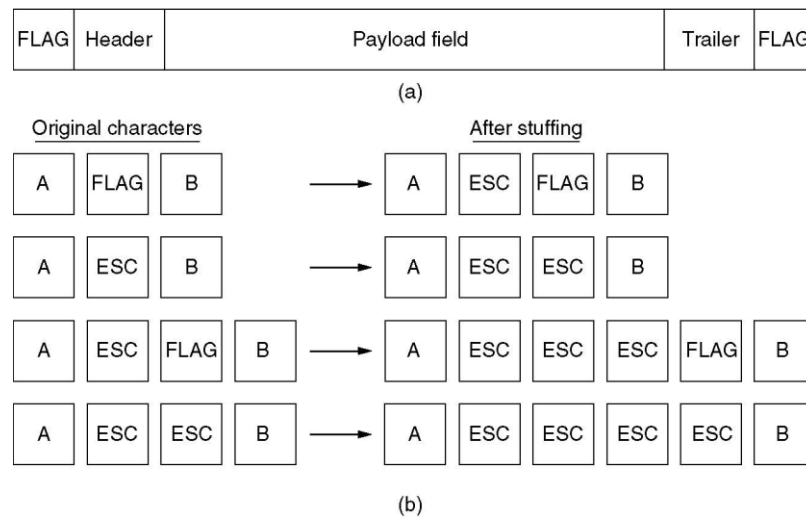


Figure 5. 8(b) Byte Stuffing and Unstuffing

iii. Starting and ending flags, with bit stuffing

This process is also known as bit oriented framing approach. Bit stuffing is the process of adding one extra 0 whenever five consecutive 1's follows a 0 in the data, so that the receiver does not mistake the pattern 0111110 for a flag. Most protocols use a special 8-bit pattern flag 01111110 as the delimiter to define the beginning and the end of the frame. This flag can create the same type of problem. That is, if the flag pattern appears in the data, we need to somehow inform the receiver that this is not the end of the frame. We do this by stuffing 1 single bit (instead of 1 byte) to prevent the pattern from looking like a flag. The strategy is called bit stuffing.

Bit Stuffing

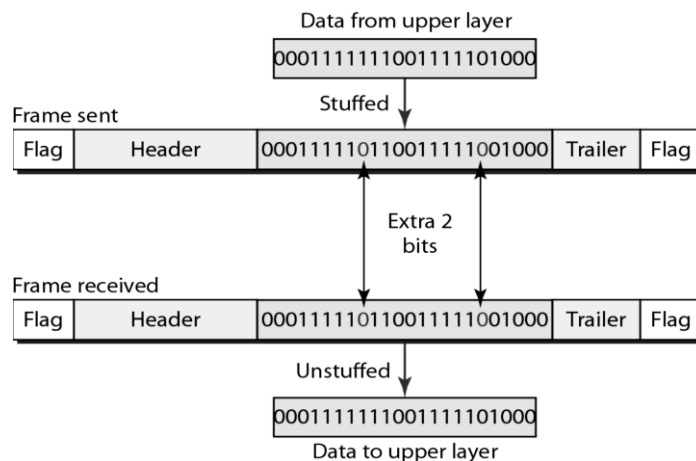


Figure 5. 9 Process of Bit Stuffing

- (a) The original data.
 (b) The data as they appear on the line.
 (c) The data as they are stored in receiver's memory after destuffing.

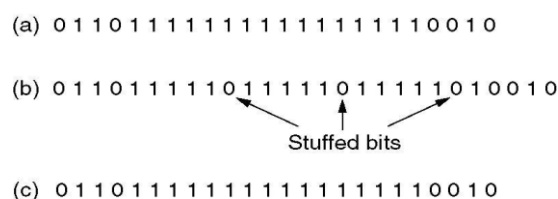


Figure 5. 10 Stuffed Bits

iv. Physical Layer Coding Violation

The last method of framing is only applicable to networks in which the encoding on the physical medium contains some redundancy

- For example, some LANs encode 1 bit of data by using 2 physical bits. Normally, a 1 bit is a high-low pair and a 0 bit is a low-high pair.
- The scheme means that every data bit has a transition in the middle, making it easy for the receiver to locate the bit boundaries.
- The combinations high-high and low-low are not used for data but are used for delimiting frames in some protocols.

As a final note on framing, many data link protocols use a combination of a character count with one of the other methods for extra safety.

When a frame arrives, the count field is used to locate the end of the frame.

- Only if the appropriate delimiter is present at that position and the checksum is correct is the frame accepted as valid.
- Otherwise, the input stream is scanned for the next delimiter.

5.2 Error Detection and Correction

- Data can be corrupted during transmission. For reliable communication, error must be detected and corrected
- Error Detection and Correction are implemented either at the data link layer or the transport layer of the OSI model.

Type of Errors

The types of errors can be understood with the help of Figure 5.10.

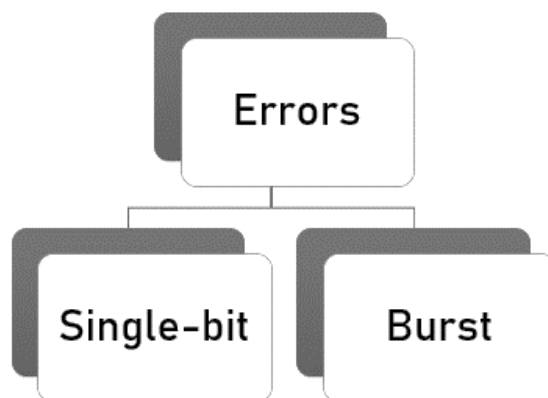


Figure 5. 11 Types of Errors

Single-Bit Error

- It is when only one bit in the data unit has changed (ex : ASCII STX - ASCII LF). This can be seen in figure 5.11.

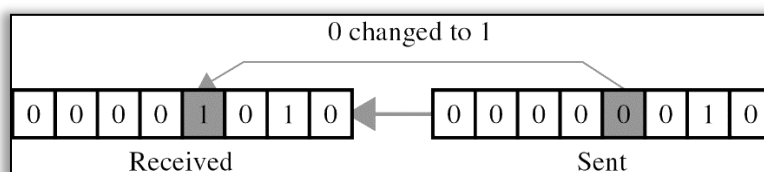


Figure 5. 12 Single Bit Errors

Multiple-Bit Error

It is when two or more nonconsecutive bits in the data unit have changed (ex : ASCII B - ASCII LF)

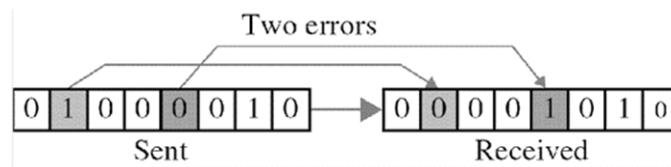


Figure 5. 13 This can be seen in figure 5.12.

Burst Error

It means that 2 or more consecutive bits in the data unit have changed. This can be seen in figure 5.13.

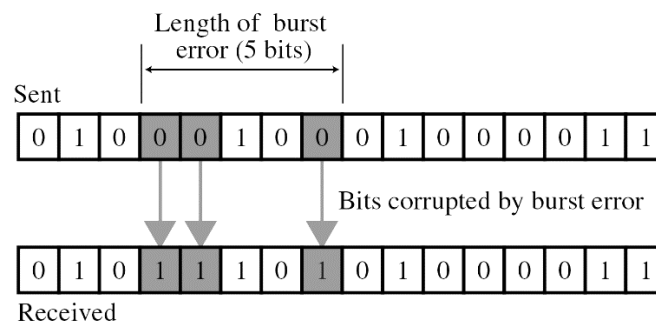


Figure 5. 14 Burst Errors

Detection

Error detection uses the concept of redundancy, which means adding extra bits for detecting errors at the destination. This can be seen in figure 5.14.

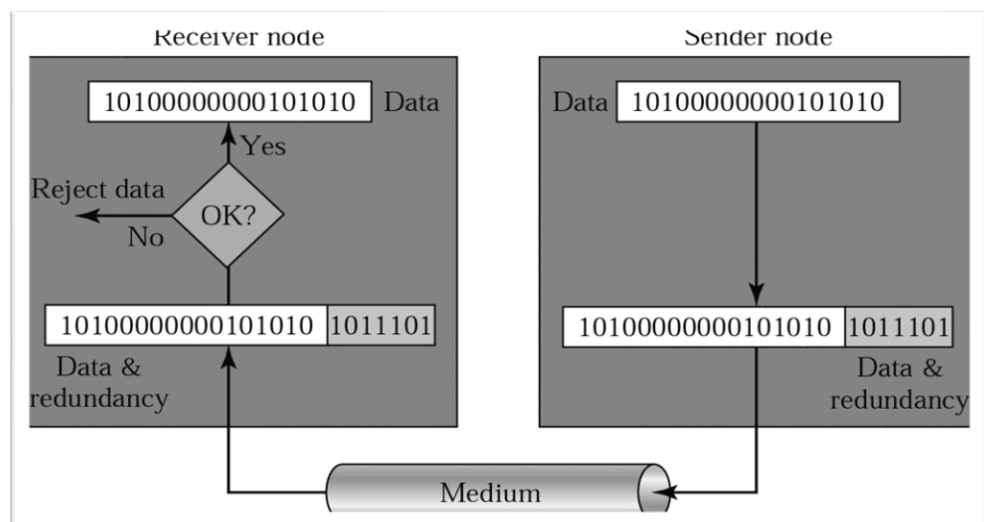


Figure 5. 15 Error Detection using Redundancy

Detection Methods

The detection methods can be categorized into three broad fields:

- Parity Check Method
- Cyclic Redundancy Check
- Checksum Method

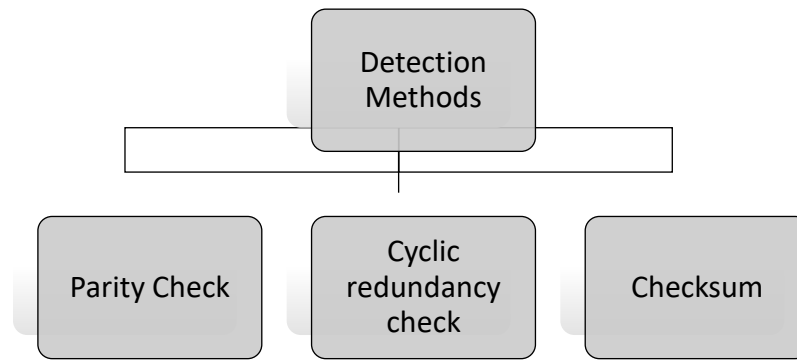


Figure 5.16 Detection Methods

a) Parity Check

A parity bit is added to every data unit so that the total number of 1s (including the parity bit) becomes even for even-parity check or odd for odd-parity check

i Simple parity check Detection

The errors can be detected in the transmitted data with the help of parity bits. Figure 5.16 shows the error detection process.

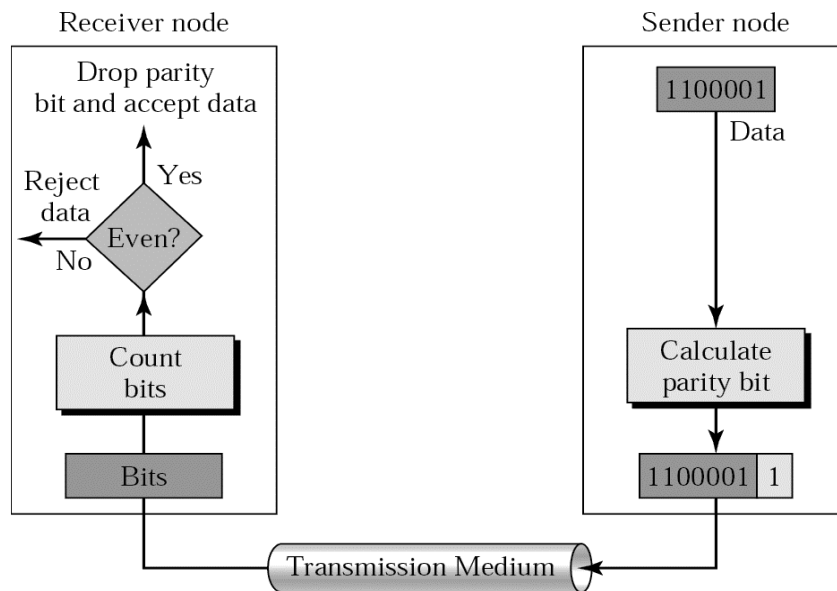


Figure 5.17 Errors detection in the transmitted data



Suppose the sender wants to send the word *world*. In ASCII the five characters are coded as

1110111 1101111 1110010 1101100 1100100

The following shows the actual bits sent

11101110 11011110 11100100 11011000 11001001



Now suppose the word *world* in Example 1 is received by the receiver without being corrupted in transmission.

11101110 11011110 11100100 11011000 11001001

The receiver counts the 1s in each character and comes up with even numbers (6, 6, 4, 4, 4). The data are accepted.



Now suppose the word world in Example 1 is corrupted during transmission.

11111110 11011110 11101100 11011000 11001001

The receiver counts the 1s in each character and comes up with even and odd numbers (7, 6, 5, 4, 4). The receiver knows that the data are corrupted, discards them, and asks for retransmission.

ii Two - Dimensional Parity Check

The two-dimensional parity checking can be understood with the help of Figure 5.17

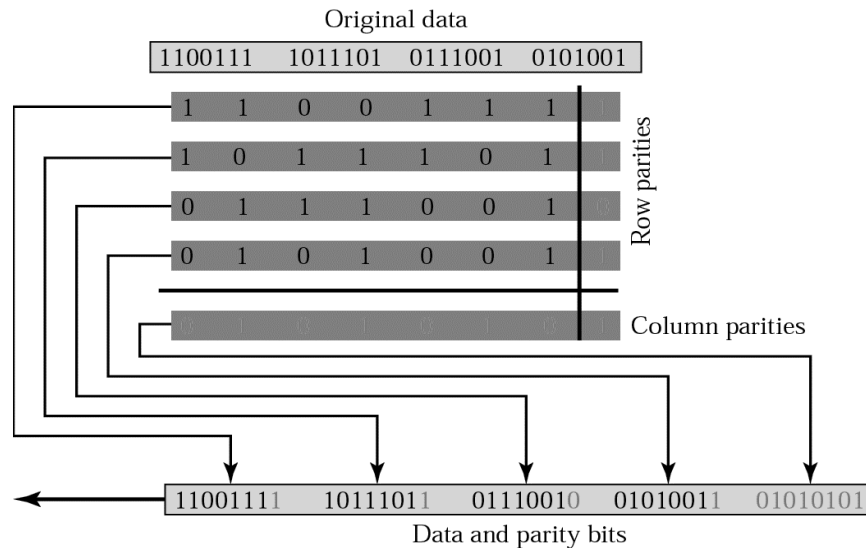


Figure 5.18 Two-Dimensional Parity Checking



Suppose the following block is sent:

10101001 00111001 11011101 11100111 10101010

However, it is hit by a burst noise of length 8, and some bits are corrupted.

10100011 10001001 11011101 11100111 10101010

10100011 10001001 11011101 11100111 10101010

b) CRC (Cyclic Redundancy Check)

It is based on binary division as is evident from Figure 5.18.

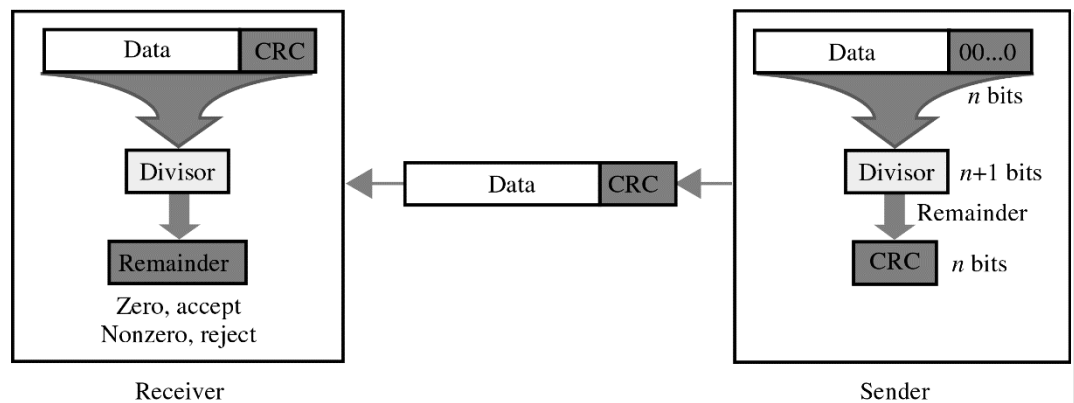


Figure 5.19 Working of Cyclic Redundancy Check

Detection with CRC generator

It uses modular-2 division.

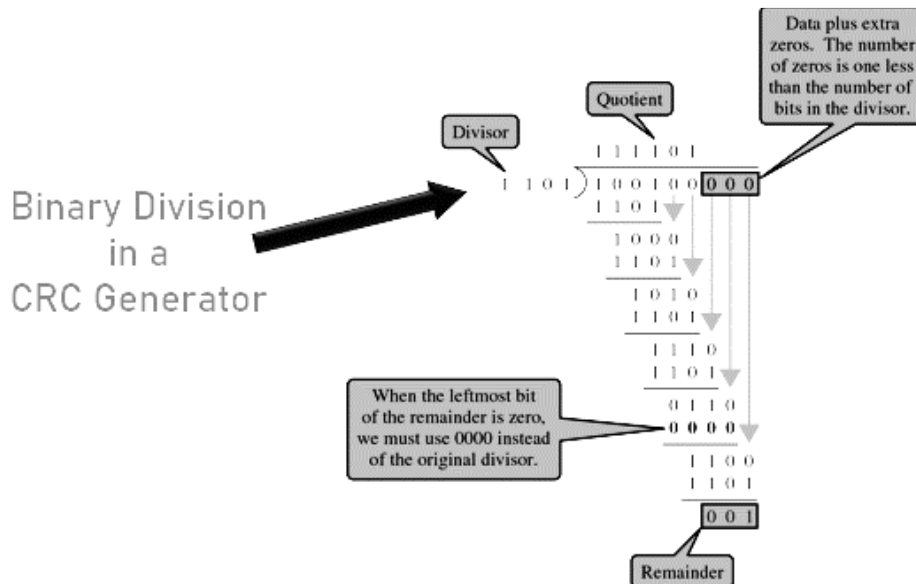


Figure 5.19 (a) Binary Division in a CRC Checker

Binary Division in a CRC Checker

To understand binary division in CRC method let's consider the following example

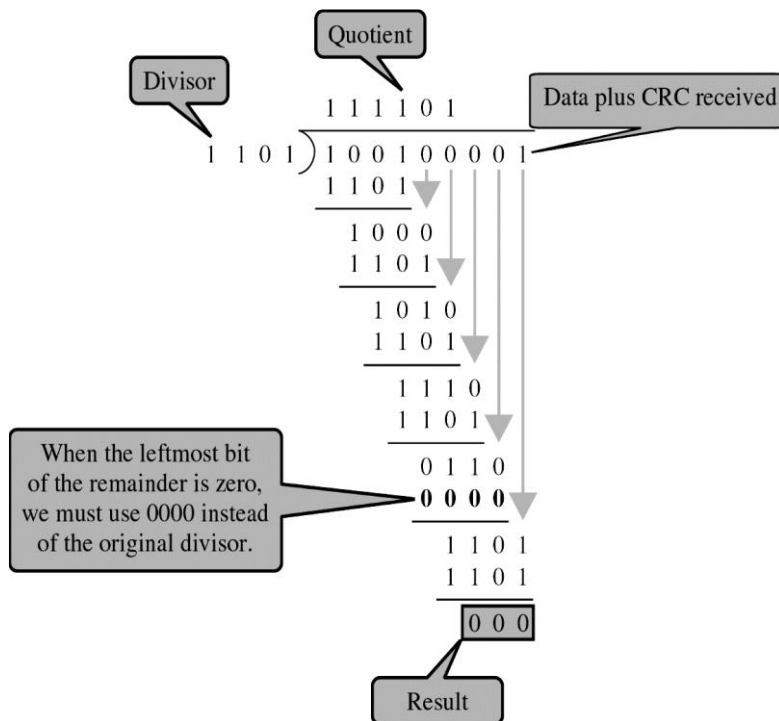


Figure 5.20 (b) Binary Division in a CRC Checker

Use of Polynomials in CRC:

The CRC generator (divisor) is most often represented not as a string of 1s and 0s, but as an algebraic polynomial. So, let's consider a polynomial expression

$$x^7 + x^5 + x^2 + x + 1$$

Now let us try to understand the polynomial representing a divisor. Let us have a look at Figure 5.20.

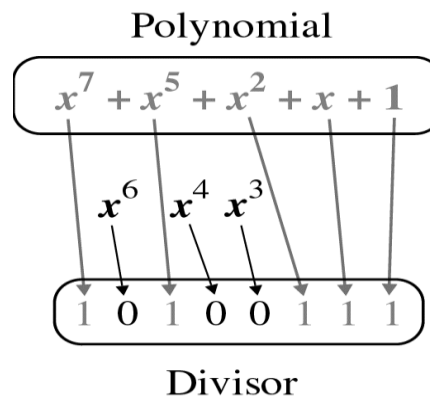


Figure 5.21 Understanding the Polynomial Representing a Divisor

Standard polynomials

Let us have a look at some of the standard polynomials shown in figure 5.21

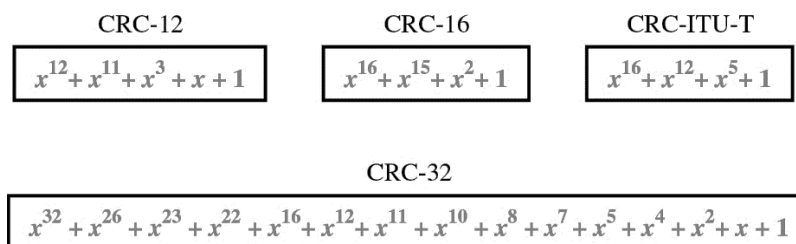


Figure 5.22 Standard Polynomials

c) Checksum

It should be noted that checksum is

- used by the higher layer protocols
- is based on the concept of redundancy (VRC, LRC, CRC)

Checksum Generator

Figure 5.22 shows the process of a checksum generator

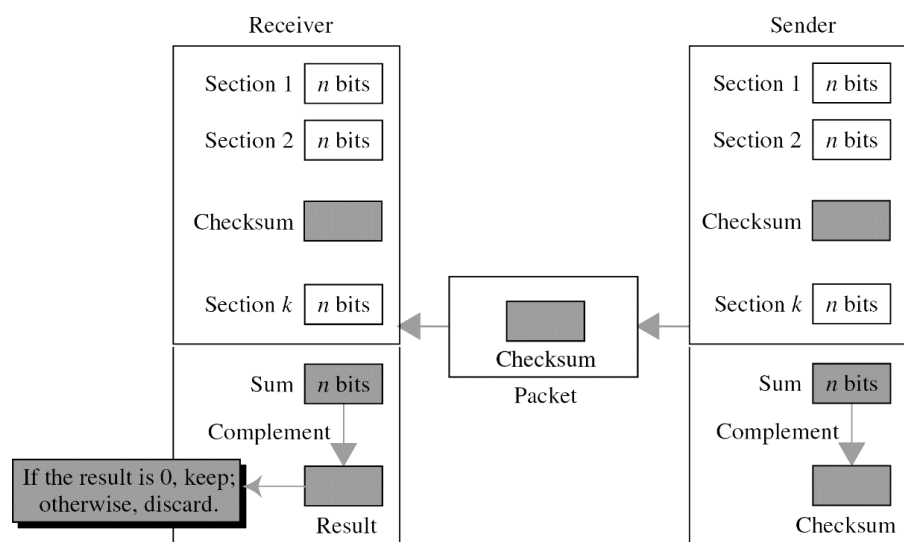


Figure 5.23 Checksum Generator

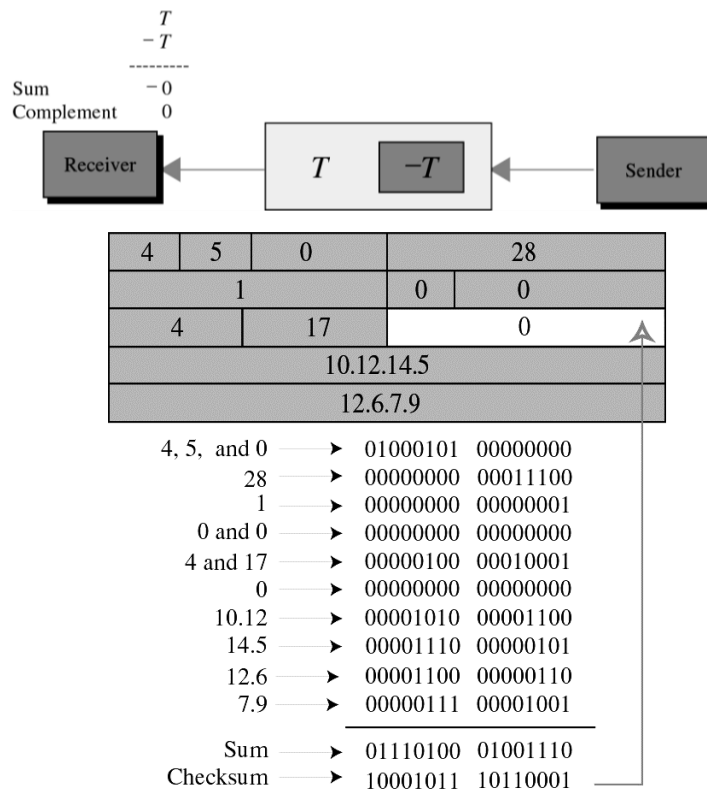
Detection

To create the checksum the sender does the following:

- The unit is divided into K sections, each of n bits.
- Section 1 and 2 are added together using one's complement.
- Section 3 is added to the result of the previous step.
- Section 4 is added to the result of the previous step.
- The process repeats until section k is added to the result of the previous step.
- The result is complemented to make the checksum.

Data unit and checksum

The receiver adds the data units and the checksum field. If the result is all 1's, the data unit is accepted; otherwise it is discarded. Let's try to understand it with the help of



(at a sender)	(at a receiver)
Original data : 10101001 00111001	Received data : 10101001 00111001 00011101
10101001	10101001
00111001	00111001
-----	00011101
11100010 Sum	-----
00011101 Checksum	11111111 ← Sum
10101001 00111001 00011101 ← Send	00000000 ← Complement

Figure 5. 24 Data Units and the Checksum

Error Correction

Errors and Error Correcting Codes

- When bits are transmitted over the computer network, they are subject to get corrupted due to interference and network problems. The corrupted bits lead to spurious data being received by the receiver and are called errors.
- Error-correcting codes (ECC) are a sequence of numbers generated by specific algorithms for detecting and removing errors in data that has been transmitted over noisy channels. Error correcting codes ascertain the exact number of bits that has been corrupted and the location of the corrupted bits, within the limitations in algorithm.

Error-correcting codes can be broadly categorized into two types –

- a) **Block codes** – The message is divided into fixed-sized blocks of bits, to which redundant bits are added for error detection or correction.
- b) **Convolutional codes** – The message comprises of data streams of arbitrary length and parity symbols are generated by the sliding application of a Boolean function to the data stream.

Errors can be handled in two ways

- when an error is discovered, the receiver can have the sender retransmit the entire data unit.
- a receiver can use an error-correcting code, which automatically corrects certain errors.

Encoding a message by Hamming Code

The procedure used by the sender to encode the message encompasses the following steps –

- Step 1 – Calculation of the number of redundant bits.
- Step 2 – Positioning the redundant bits.
- Step 3 – Calculating the values of each redundant bit.
- Once the redundant bits are embedded within the message, this is sent to the user.

Step 1 – Calculation of the number of redundant bits.

- If the message contains m number of data bits, r number of redundant bits are added to it so that $m+r$ is able to indicate at least $(m+r+1)$ different states. Here, $(m+r)$ indicates location of an error in each of $(m+r)$ bit positions and one additional state indicates no error. Since, r bits can indicate 2^r states, 2^r must be at least equal to $(m+r+1)$. Thus the following equation should hold $2^r \geq m+r+1$

Step 2 – Positioning the redundant bits.

- The r redundant bits placed at bit positions of powers of 2, i.e. 1, 2, 4, 8, 16 etc. They are referred in the rest of this text as r_1 (at position 1), r_2 (at position 2), r_3 (at position 4), r_4 (at position 8) and so on.

Step 3 – Calculating the values of each redundant bit.

- The redundant bits are parity bits. A parity bit is an extra bit that makes the number of 1s either even or odd. The two types of parity are –
- Even Parity – Here the total number of bits in the message is made even.
- Odd Parity – Here the total number of bits in the message is made odd.
- Each redundant bit, r_i , is calculated as the parity, generally even parity, based upon its bit position. It covers all bit positions whose binary representation includes a 1 in the i th position except the position of r_i .

Thus

- r_1 is the parity bit for all data bits in positions whose binary representation includes a 1 in the least significant position excluding 1 (3, 5, 7, 9, 11 and so on)
- r_2 is the parity bit for all data bits in positions whose binary representation includes a 1 in the position 2 from right except 2 (3, 6, 7, 10, 11 and so on)

- r_3 is the parity bit for all data bits in positions whose binary representation includes a 1 in the position 3 from right except 4 (5-7, 12-15, 20-23 and so on)

Decoding a message in Hamming Code

- Once the receiver gets an incoming message, it performs recalculations to detect errors and correct them. The steps for recalculation are –
- Step 1 – Calculation of the number of redundant bits.
- Step 2 – Positioning the redundant bits.
- Step 3 – Parity checking.
- Step 4 – Error detection and correction

Single-Bit Error Correction

- Parity bit
- The secret of error correction is to locate the invalid bit or bits
- For ASCII code, it needs a three-bit redundancy code (000-111)

Error Correction

- Redundancy Bits: To calculate the number of redundancy bits (R) required to correct a given number of data bit (M). It is depicted in figure 5.24.

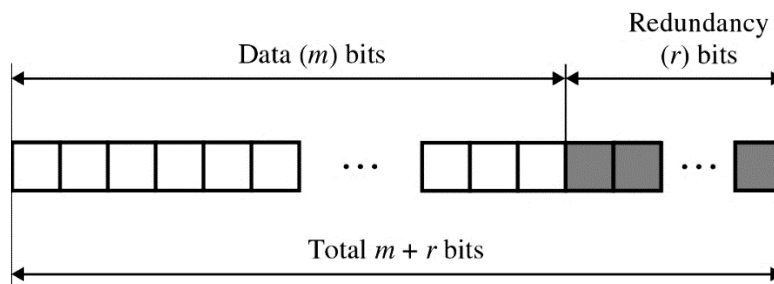


Figure 5. 25 Position of Redundant Bits

- If the total number of bits in a transmittable unit is $m+r$, then r must be able to indicate at least $m+r+1$ different states

$$2^r \geq m + r + 1$$

Example: For value of m is 7 (ASCII), the smallest r value that can satisfy this equation is 4

$$2^4 \geq 7 + 4 + 1$$

Table 5. 1 Relationship between Data and Redundant Bits

Number of Data Bits (m)	Number of Redundancy Bits (r)	Total Bits ($m+r$)
1	2	3
2	3	5
3	3	6
4	3	7
5	4	9
6	4	10
7	4	11

To understand the relationship between data and redundancy bits, have a look at Table 5.1

Hamming Code

Hamming code is a block code that is capable of detecting up to two simultaneous bit errors and correcting single-bit errors. It was developed by R.W. Hamming for error correction. In this coding method, the source encodes the message by inserting redundant bits within the message. These redundant bits are extra bits that are generated and inserted at specific positions in the message itself to enable error detection and correction. When the destination receives this message, it performs recalculations to detect errors and find the bit position that has error.

It was developed by R. W. Hamming

- The figure 5.25 shows the positions of redundancy bits in Hamming code
- The redundancy bits will be at positions $2^4, 2^2, 2^1, 2^0$

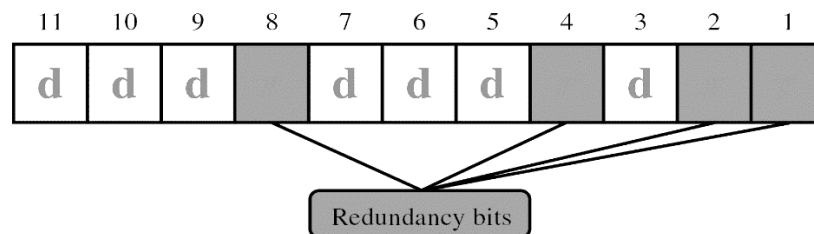


Figure 5. 26 The Positions of the Redundant Bits

Each r bit is the VRC bit for one combination of data bits: -

r_1 = bits 1, 3, 5, 7, 9, 11

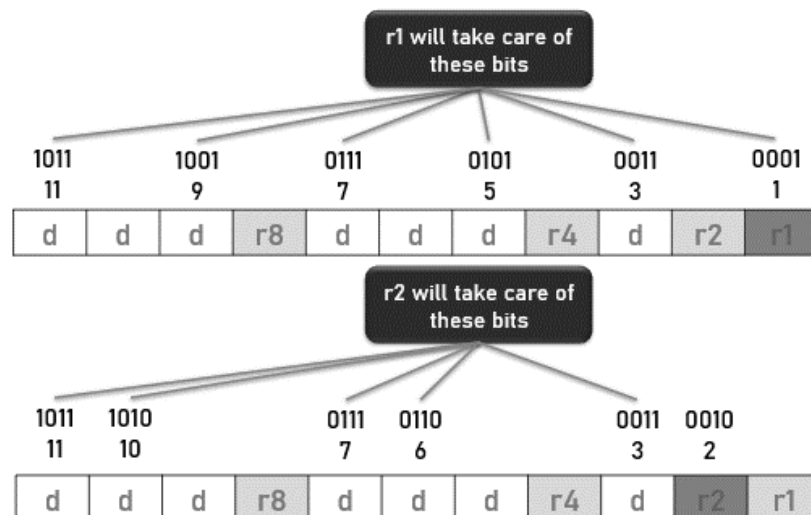
r_2 = bits 2, 3, 6, 7, 10, 11

r_4 = bits 4, 5, 6, 7

r_8 = bits 8, 9, 10, 11

Redundancy bits calculation

To understand the procedure of redundant bits calculation, let's have a look at figure 5.26



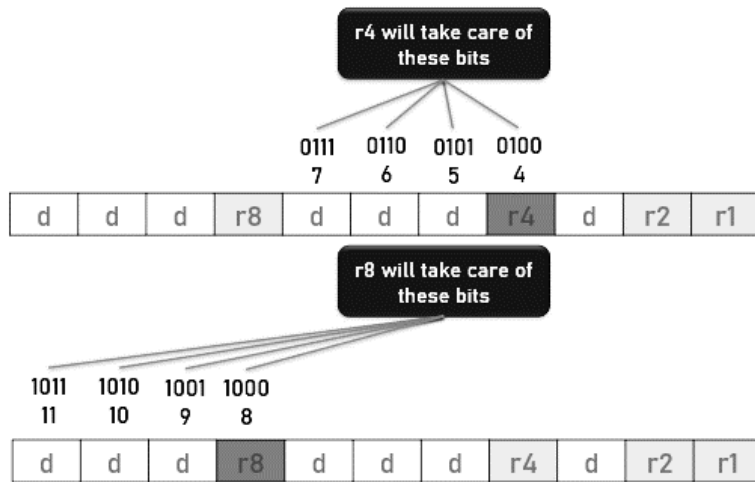


Figure 5.27 Calculating the Redundant Bit Values

Now let's understand how to find errors with the help of redundant bits with figure 5.27. it is finding errors by calculating the even parity

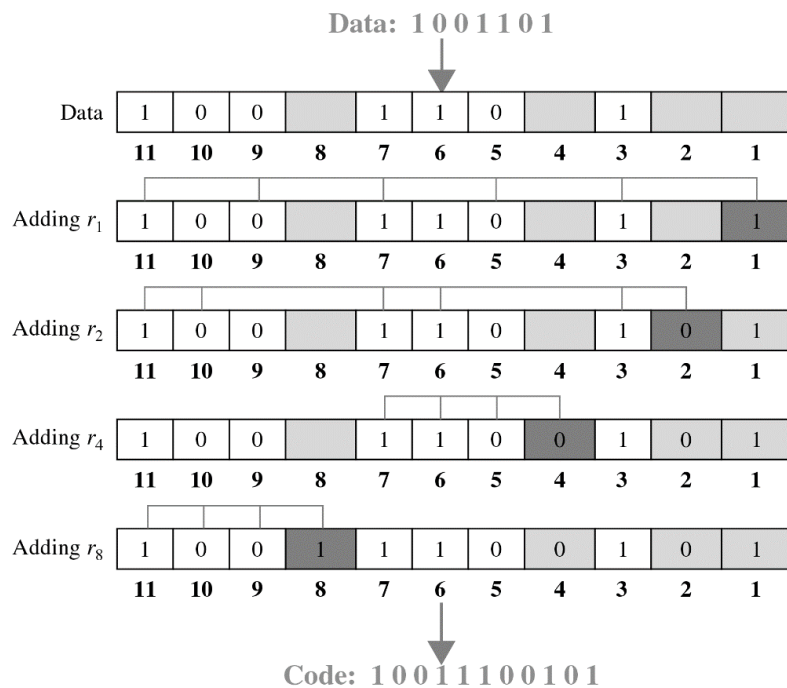
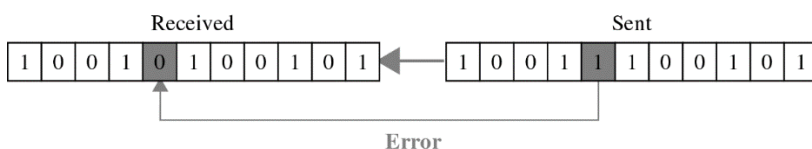


Figure 5.28 Finding errors by calculating the even parity

Error Detection and Correction

Now to see how errors are detected and corrected have a look at figure 5.28. it can be clearly seen that the bit positions value is changing from 1 to 0. When the parity of the different between sessions is calculated we can see that the bit position 7 is having an error because the parity positions have a final value 0 1 1 1 which in decimal is equal to 7. This clearly depicts that the error is at bit position 7.



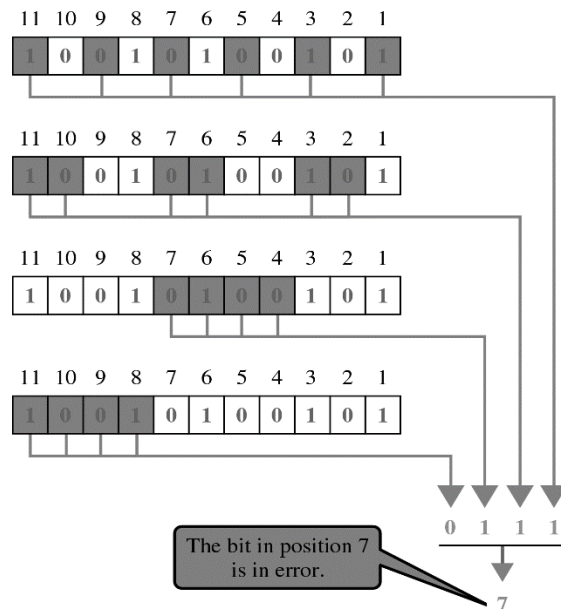


Figure 5. 29 Detecting Errors with the help of Parity Bit Positions

Multiple-Bit Error Correction

Redundancy bits calculated on overlapping sets of data units can also be used to correct multiple-bit errors.

Example:

To correct double-bit errors, we must take into consideration that two bits can be a combination of any two bits in the entire sequence.

Hamming Code for Error Detection & Correction

Hamming code can be used to detect and correct errors. For example, Whenever a sender sends a message, if the receiver doesn't receive the same exact message as the sender had sent, this means that some error is there.

This error can be found with the help of Hamming Code and we can correct the errors also.

- It can be applied to data units of any length
- It is used to detect and correct the single bit errors

Now let us see the Hamming Code structure.

All the bit positions that are power of 2 are marked as parity bits i.e. ($2^0=1, 2^1=2, 2^2=4, 2^3=8, \dots$). All other bits are for data

D7	D6	D5	<u>P4</u>	D3	<u>P2</u>	<u>P1</u>
----	----	----	-----------	----	-----------	-----------

Now let us determine the value of the parity bits. We already know about the Data bits given by the user, but we don't know about the Parity bit values. So how do we calculate the value of the parity bits. Let us understand the rules

Rule:

The value of the parity bit is determined by the sequence of bits. It follows alternative Checks and Skips.

Let us understand this with the help of an example.

1 1 0 1

Sender----->Receiver

D7	D6	D5	P4	D3	P2	P1
1	1	0		1		

D7	D6	D5	P4	D3	P2	P1
1	1	0		1		

For P1 (Check 1 bit - Skip 1 Bit)

1 ~~2~~ 3 ~~4~~ 5 ~~6~~ 7 ~~8~~ 9

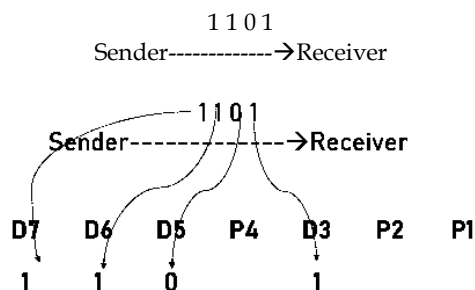
For P2 (Check 2 bits - Skip 2 bits)

2 3 ~~4~~ ~~5~~ 6 7 ~~8~~ ~~9~~ 10 11

For P4 (Check 4 bits - Skip 4 bits)

4 5 6 7 ~~8~~ ~~9~~ ~~10~~ ~~11~~ 12 13 14 15 ~~16~~ ~~17~~ ~~18~~ ~~19~~ 20 21 22 23

P1				P2				P4			
P1	D3	D5	D7	P2	D3	D6	D7	P4	D5	D6	D7



P1			
P1	1	0	1

Since the number of 1's is even, so set P1 = 0

P2			
P2	1	1	1

Since the number of 1's is odd, so set P2 = 1

P4			
P4	0	1	1

Since the number of 1's is even, so set P4 = 0

So, the Parity bits are

P1				P2				P4			
0	1	0	1	1	1	1	1	0	0	1	1

and

D7	D6	D5	P4	D3	P2	P1
1	1	0	0	1	1	0

Now let us understand the process of Detecting errors in Hamming code structure. So, let us consider a 7 Bit Hamming code structure

D7 D6 D5 P4 D3 P2 P1

So let us examine the receiver's end bits. If the parity is not even for any group of bits, it means no errors.

Correction of errors:

The decimal equivalent of the parity bits binary values is calculated. If it is 0, there is no error. Otherwise, the decimal value gives the bit position which has error. For example, if c1c2c3c4 = 1001, it implies that the data bit at position 9, decimal equivalent of 1001, has error. The bit is flipped to get the correct message.

Summary

Data link layer describes the techniques to access a shared communication channel and reliable data transmission. Its main tasks are framing, checksums, error detection and correction, acknowledgement, flow control, encapsulating packets from network layer to frames, etc. The data link layer provides unacknowledged connectionless service, acknowledged connectionless service and acknowledged connection-oriented service. Parity check is the simplest form of error detection method as the receiver needs to count only the number of 1's in the received data stream with additional parity bit. Checksum is a simple type of redundancy check used to detect errors in data. Cyclic Redundancy Check is used widely in computer networks, is a technique of providing a data string added to packets of information that can be used to detect errors in the data packets. Stop and Wait protocol is easiest to implement and proves to be the most efficient on an error free communication channel. However, an error free communication channel is practically not possible.

Keywords

Acknowledged Connection-oriented Service: The data link layer provides this service to the network layer by establishing a connection between the source and destination hosts before any transfer of data takes place.

Acknowledged Connectionless Service: Refers to delivery of each frame sent between two hosts arrives correctly.

Checksum: Refers to an algorithm that calculates the binary values in a packet or other block of data and stores the results with the data to compare with a new checksum at the other end.

Cyclic Redundancy Check: Refers to a technique of providing a data string added to packets of information that can be used to detect errors in the data packets.

Error Control: Involves sequencing frames and sending control frames for acknowledgement.

Flow Control: Refers to control the rate of data transmission between two source and destination hosts.

Framing: Provides a reliable transfer of bit streams to the network layer the data link layer breaks the bit stream into frames.

Go Back N: The Go Back N protocol enables the source machine to have more than one outstanding frame at a time by using buffers.

High-level Data Link Control: Refers to receipt of data that is checked after multiple frames are sent for improved transmission efficiency. It also offers a form of advanced error control called CRC (Cyclic Redundancy Check).

Parity Checks: Consists of even parity and odd parity method. The operation of receiver is simple, as the receiver needs to count only the number of 1's in the received data stream with additional parity bit.

Unacknowledged Connectionless Service: Refers to the independent frames from source host to the destination host without any acknowledgment mechanism.

Self Assessment

Fill in the blanks:

1. The data link layer receives a raw bit stream from the layer that may not be error free.
2. Some of the examples of check are audio storage and playback devices such as audio CD's.
3. CRC codes are also called as codes.
4. consists of Even Parity and Odd Parity Method.

State whether the following statements are true or false:

5. Shannon's theorem is an important theorem in forward error correction.
6. The actual maximum code rate allowed depends on the error-correcting code used.
7. The code rate is defined as the fraction k/n of k source symbols and n encoded symbols.
8. Block codes are processed on a bit-by-bit basis.
9. Early examples of block codes are repetition codes, hamming codes and multidimensional parity-check codes.
10. Turbo codes and low-density parity-check codes (ldpc) are relatively new constructions that can provide almost optimal efficiency.

Select the correct answer for the following questions

11. Which of the following is not a function performed by the Data Link Layer?
 - a) Reliable data transfer service between two peer network layers
 - b) It provides a logical communication between application processes running on different hosts.
 - c) Flow Control mechanism which regulates the flow of frames to avoid data congestion
 - d) It encapsulates the received packets into Frames.
12. Which of the following is true regarding Error Control?
 - a) Sometimes signals may have encountered problem in transition and the bits are flipped.
 - b) It involves ensuring both machine to exchange data on same speed.

- c) The errors are detected and attempted to recover actual data bits.
 - d) It provides error reporting mechanism to the sender.
13. Which of the following is true regarding the Logical Link Control (LLC)?
- a) It explains how to share the link.
 - b) It deals with the design and procedures for communication b/w nodes.
 - c) It provides mechanism such as CSMA/CD to equip capability of accessing shared media among multiple Systems.
 - d) All the mentioned choices
14. The function that the Logical link control performs includes_____.
- a) Framing
 - b) Flow Control
 - c) Error control
 - d) All the given choices
15. Which of the following is not true regarding Multiple-Access?
- a) When host on the shared link tries to transfer the data, it has a high probability of collision.
 - b) Data-link layer provides CSMA/CD to equip accessing a shared media among multiple Systems.
 - c) It provides error reporting mechanism to the sender.
 - d) None of the given choices

Review Questions

1. What is the data link protocol?
2. What advantages does Selective Repeat sliding window protocol offer over Go Back N protocol?
3. What is the purpose of flow control?
4. Describe how does finite state machine model carry out protocol verification?
5. What are different data link protocols available? Why does PPP have become popular?

Answers: Self Assessment

- | | | | |
|-------------|-----------|---------------|-----------------|
| 1. physical | 2. Parity | 3. Polynomial | 4. Parity check |
| 5. True | 6. True | 7. True | 8. False |
| 9. True | 10. True | 11. b | 12. b |
| 13. b | 14. d | 15. c | |

Further Readings



Andrew S. Tanenbaum, *Computer Networks*, Prentice Hall.

Behrouz A. Forouzan and Sophia Chung Fegan, *Data Communications and Networking*, McGraw-Hill Companies.

J. D. Spragins, *Telecommunications Protocols and Design*, Addison-Wesley.



<https://www.geeksforgeeks.org/hamming-code-in-computer-network/>

Unit 06: Data Link layer - Flow and Error Control Protocols

CONTENTS

Objectives

Introduction

6.1 Error Control

6.2 Flow Control

6.3 Data Link Control Protocols

6.4 Noisy Channels

Keywords

Self Assessment

Review Questions

Answers: Self Assessment

Further Readings

Objectives

After this lecture, you would be able to

- understand about the basic concepts of flow control
- learn the basic protocols for noisy and noiseless channels
- learn the sliding window protocol

Introduction

The data link layer is one of the most important layers of the OSI Reference model. One of the most critical functions performed by the data link layer is the transference of data in a network. However, this requires different mechanism when transferring data in a noisy or a noiseless channel. So different rules or protocols are defined to organize this flow of data from the source machine to the destination machine. The prominent protocols for the noiseless channels are the Simplest Protocol and the Stop-and-Wait Protocol. In case of the noisy channel there are three prominent protocols which are the Stop-and-Wait Automatic Repeat Request, Go-back-N Automatic Repeat Request and the Selective Repeat Automatic Repeat Request.

6.1 Error Control

When data-frame is transmitted, there is a probability that data-frame may be lost in the transit or it is received corrupted. In both cases, the receiver does not receive the correct data-frame and sender does not know anything about any loss. In such case, both sender and receiver are equipped with some protocols which helps them to detect transit errors such as loss of data-frame. Hence, either the sender retransmits the data-frame or the receiver may request to resend the previous data-frame.

Requirements for error control mechanism:

The various requirements of the error control mechanisms are:

- ***Tracing the Error:*** Either the sender and/or the receiver, must ascertain that there is some error in the transit.
- ***A positive-ACK:*** is generated when the receiver receives a correct frame.
- ***A negative-ACK:*** When the receiver receives a damaged frame or a duplicate frame, it sends a NACK back to the sender and the sender must retransmit the correct frame.

- **The retransmission:** The sender maintains a clock and sets a timeout period. If an acknowledgement of a data-frame previously transmitted does not arrive before the timeout the sender retransmits the frame, thinking that the frame or it's acknowledgement is lost in transit.

6.2 Flow Control

The data link layer is mainly responsible for the proper flow of data from the sender to the receiver. This becomes a lot more challenging in case the receiver has high traffic load and there is a mismatch in the speed of sending data by the sender and the speed of receiving data by the receiver. This can lead to loss frames which is highly undesirable. Problem is not only about the loss of frames but is also about the delete acknowledgements being sent by the receiver in both the cases the frame will have to be retransmitted. Not the main question that arises is how should a receiver signal a sender that it has the peak load and the sender needs to slow down the speed of transmission of the frames. To do so the data link layer uses various protocols for the noisy as well as the noiseless channels.

Flow control refers to a set of procedures used to restrict the amount of data that the sender can send before waiting for acknowledgment.

Two types of mechanisms can be deployed to control the flow:

- **Stop and Wait:** This flow control mechanism forces the sender after transmitting a data frame to stop and wait until the acknowledgement of the data-frame sent is received.
- **Sliding Window:** In this flow control mechanism, both sender and receiver agree on the number of data-frames after which the acknowledgement should be sent. As we learnt, stop and wait flow control mechanism wastes resources, this protocol tries to make use of underlying resources as much as possible.

6.3 Data Link Control Protocols

The data link layer takes the responsibility of transferring data from the sender to the receiver. However, this becomes challenging due to varying type of channels. Since a network can have a noisy or a noiseless channel. So, the data link layer needs to have different protocols for these different channels and their needs. Let's have a look at the various protocols used by the data link layer for data transmission. Figure 6.1 shows the different protocols which can be used for both the types of channels.

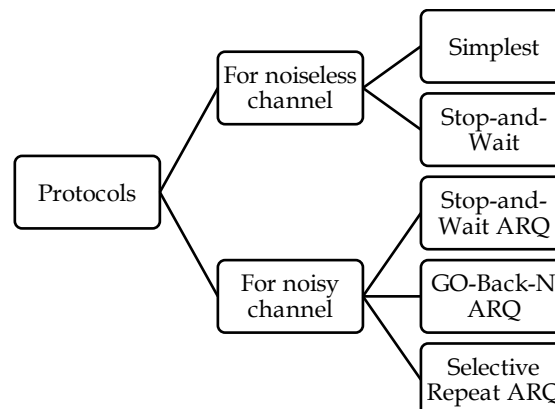


Figure 6.1 Protocols for Noisy and Noiseless Channels

Now let us discuss the different types of protocols in the Noisy and the Noiseless channels

Noiseless Channel

A Noiseless channel is an ideal channel in which no frames are lost, duplicated, or corrupted. It has two protocols as can be seen in figure 6.2:

- The first protocol does not use flow control
- The second uses flow control

Of course, neither has error control because the channel is a perfect noiseless channel.

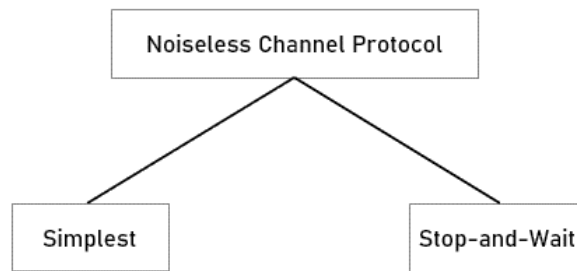


Figure 6.2 Noiseless Channel Protocol

a) Simplest Protocol

Simplest Protocol has no flow control and error control mechanism. Here the data frames travel in only one direction (from sender to receiver). The receiver can immediately handle any received frame with a processing time that is negligible. We assume that the receiver can immediately handle any frame it receives with a processing time that is small enough to be negligible. The data link layer of the receiver immediately removes the header from the frame and hands the data packet to its network layer, which can also accept the packet immediately. The receiver can never be overwhelmed with incoming frames. The protocol consists of two distinct procedures:

i) Sender

ii) Receiver

The sender and receiver run in the data link layer of the source and the destination machine respectively. No sequence number or acknowledgements are used here.

Procedure used by both data link layers:

While sending the frames, the sender site cannot send a frame until its network layer has a data packet to send. The receiver site cannot deliver a data packet to its network layer until a frame arrives. If the protocol is implemented as a procedure, we need to introduce the idea of events in the protocol. The procedure at the sender site is constantly running; there is no action until there is a request from the network layer. The procedure at the receiver site is also constantly running, but there is no action until notification from the physical layer arrives. Both procedures are constantly running because they do not know when the corresponding events will occur.

Sender-Site Algorithm for the Simplest Protocol

```

1  while (true)                                // Repeat forever
2  {
3      WaitForEvent()                          // Sleep until an event occurs
4      if( Event( RequestToSend ) )           // There is a packet to send
5      {
6          GetData();
7          MakeFrame();
8          SendFrame();                        // Send the frame
9      }
10 }
```

Figure 6.3 Sender-site Algorithm for the Simplest Protocol

The algorithm as can be seen in figure 6.3, has an infinite loop, which means the third line till the ninth line are repeated forever in a loop once the program starts. It is an event-driven algorithm, which means that it sleeps (third line) until an event wakes it up (fourth line). So, there may be an undefined gap of time between the execution of the third and the fourth line. When the event, a request from the network layer, occurs, lines numbers from six to eight are executed. The program then repeats the loop and again sleeps at the third line until the next occurrence of the event.

- GetData() its's main responsibility is to fetch a data packet from the network-layer.

- MakeFrame() It's main responsibility is adding a header and delimiter flags to the data packet for making the frame.
- SendFrame() is mainly responsible for delivering the frame to the physical layer which can further transmit it.

Receiver-site Algorithm for the Simplest Protocol

The various steps performed by the receiver-site algorithm can be seen in Figure 6.4.

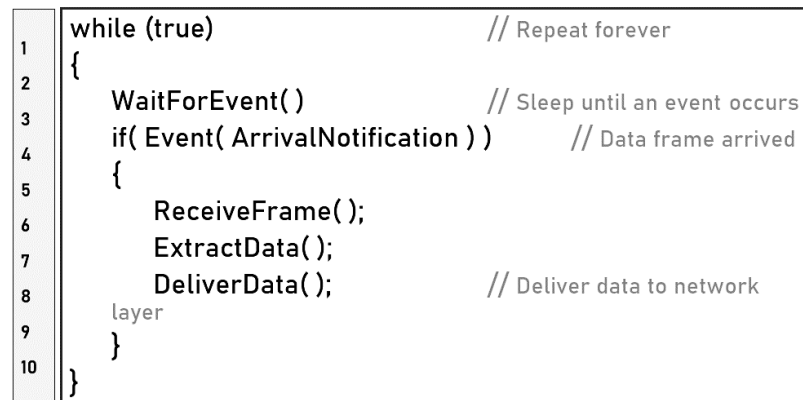


Figure 6.4 Receiver-site Algorithm for the Simplest Protocol

The event here is the arrival of a data frame. After the event,

- The data link layer receives the frame from the physical layer using the ReceiveFrame() process,
- It extracts the data from the frame using the ExtractData() process, and
- It delivers the data to the network layer using the DeliverData() process.

We have an event-driven algorithm because the algorithm never knows when the data frame will arrive.

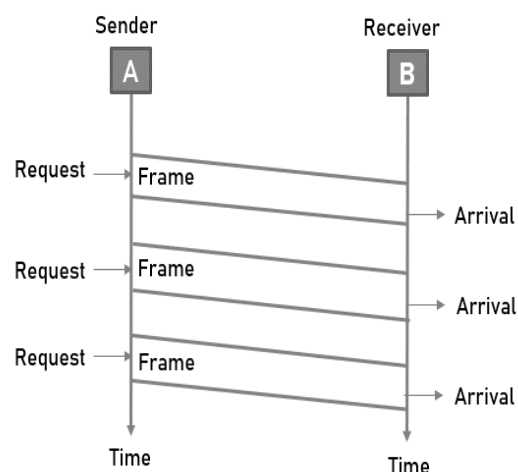


Figure 6.5 Flow Diagram

Figure 6.5 shows an example of communication using this protocol. The sender sends a sequence of frames without even thinking about the receiver. To send three frames, three events occur at the sender site and three events at the receiver site.



The data frames are shown by tilted boxes; the height of the box defines the transmission time difference between the first bit and the last bit in the frame.

b) Stop and Wait Protocol in Noiseless Channels

In a network if the data frames arrive at the receiver site faster than they can be processed, the frames must be stored until their use. Normally, the receiver does not have enough storage space, especially if it is receiving data from many sources. This may result in either the discarding of frames or denial of service. To prevent the receiver from becoming overwhelmed with frames, we somehow need to tell the sender to slow down. There must be feedback from the receiver to the sender. The sender sends one frame, stops until it receives confirmation from the receiver, and then sends the next frame. We have unidirectional communication for data frames, but auxiliary ACK frames travel from the other direction. We add flow control to our previous protocol. We can see the traffic on the forward channel and the reverse channel. At any time, there is either one data frame on the forward channel or one ACK frame on the reverse channel. We need a half-duplex link. The sender sends one frame, stops until it receives confirmation from the receiver, and then sends the next frame. We have unidirectional communication for data frames, but auxiliary ACK frames travel from the other direction. We add flow control to our previous protocol. We can see the traffic on the forward channel and the reverse channel. At any time, there is either one data frame on the forward channel or one ACK frame on the reverse channel. We need a half-duplex link. Let us try to understand this with the help of the figure 6.6.

Here the transmitter (Station A) sends a frame over the communication line and then waits for a positive or negative acknowledgement from the Receiver (station B). If no error occurs, station B sends a positive acknowledgement (ACK) to station A. The transmitter starts to send the next frame.

If frame is received at station B with errors, then a negative acknowledgement (NAK) is sent to station A. Then the station 'A' must retransmit the old packet in a new frame. There is also a possibility that the information frames or ACKs may get lost. The sender is equipped with a timer. If no recognizable acknowledgement is received and the time out interval takes place, the same frame is sent again. The sender which sends one frame and then waits for an acknowledgement before process is known as stop and wait. Figure 6.7 shows the working of the Stop-and-Wait protocol.

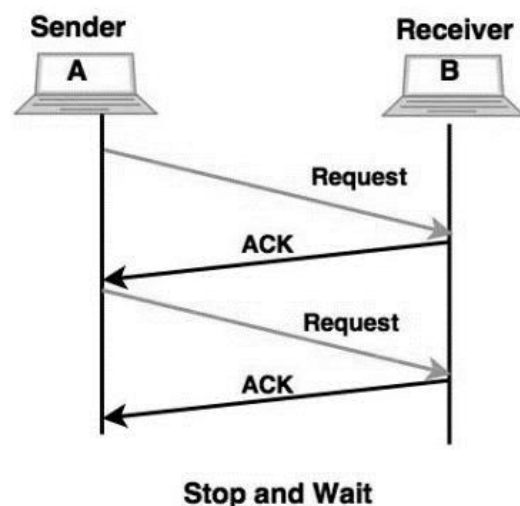


Figure 6. 6 The Stop and Wait Protocol

Stop-and-Wait Protocol Server-site Algorithm

```

1  canSend = true           //Allow the first frame to go
2  while(true)              //Repeat forever
3  {
4      WaitForEvent()i      // Sleep until an event occurs
5      if(Event(RequestToSend) AND canSend)
6      {
7          GetData();
8          MakeFrame();
9          SendFrame()i      //Send the data frame
10         canSend = false;  //cannot send until ACK arrives
11     }
12     WaitForEvent()i      // Sleep until an event occurs
13     if(Event(ArrivalNotification) // An ACK has arrived
14     {
15         ReceiveFrame();    //Receive the ACK frame
16         canSend = true;
17     }
18 }

```

Figure 6. 7 Server-Side Algorithm

Here two events can occur:

- A request from the network layer or
- An arrival notification from the physical layer

As we can see in the Figure 6.8 given below, the responses to these events must alternate. After a frame is sent, the algorithm must ignore another network layer request until that frame is acknowledged. We know that two arrival events cannot happen one after another because the channel is error-free and does not duplicate the frames. The requests from the network layer, however, may happen one after another without an arrival event in between. We need somehow to prevent the immediate sending of the data frame. Although there are several methods, we have used a simple `canSend` variable that can either be true or false. When a frame is sent, the variable is set to false to indicate that a new network request cannot be sent until `canSend` is true. When an ACK is received, `canSend` is set to true to allow the sending of the next frame.

```

1  while(true)              // Repeat forever
2  {
3      WaitForEvent( );      // Sleep until an event
4      occurs
5      if( Event( ArrivalNotification)) //Data frame arrives
6      {
7          ReceiveFrame( );
8          ExtractData( )i
9          DeliverData( );    // Deliver data to
10         network layer
11         SendFrame( );      //Send an ACK
12         frame
13     }
14 }

```

Figure 6. 8 The Stop-and-Wait Protocol Receiver-site Algorithm

After the data frame arrives, the receiver sends an ACK frame (line 9) to acknowledge the receipt and allows the sender to send the next frame. The sender sends one frame and waits for feedback from the receiver. When the ACK arrives, the sender sends the next frame.

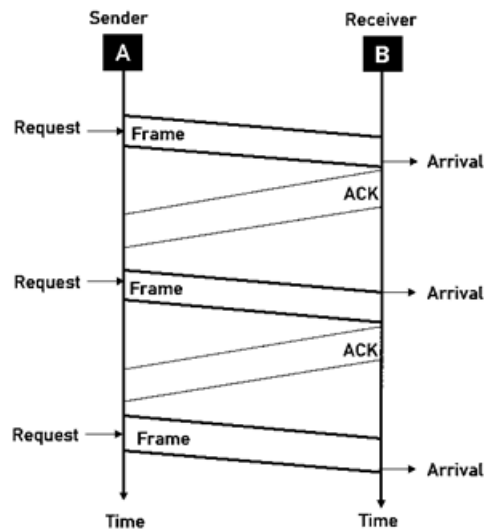


Figure 6.9 Sender and Receiver Communication Process

The sender sends one frame and waits for feedback from the receiver. When the ACK arrives, the sender sends the next frame.

Simplest Protocol (Noiseless channel) has no flow or error control. Stop-and-Wait Protocol (Noiseless channel) gives an idea of how to add flow control to Simplest Protocol. Both of these protocols do not have a mechanism to control error. Noiseless channels are nonexistent.

- Either we will be ignoring the errors, or
- We need to add error control to our protocols.
- Noisy Protocols use error control.

6.4 Noisy Channels

Consider the normal situation of a communication channel that makes errors. Frames may be either damaged or lost completely.

Data Link Control

- Physical layer is only responsible for data transmission
- Data link control is responsible for converting transmission to communication
 - Line Discipline:** Who should send now?
 - Flow Control:** How much data may be sent?
 - Error Control:** How should the errors be detected/ corrected?

i) Line Discipline

It determines the direction of communication. It makes sure that receiver is ready to accept or signal the sender to start. To do this there are two ways:

- Enquiry / Acknowledgment (ENQ/ACK) Dedicated line between hosts
- Poll / Select Multipoint connections

Line Discipline (ENQ/ACK)

It has a dedicated line between hosts, no problem of addressing. It Coordinates which device may start transmission, and if the receiver is ready and enabled. If both hosts have equal ranks, either can initiate the process. Otherwise, only higher-ranked host is allowed to start the transmission request. Can be run in either half-duplex or full-duplex modes.

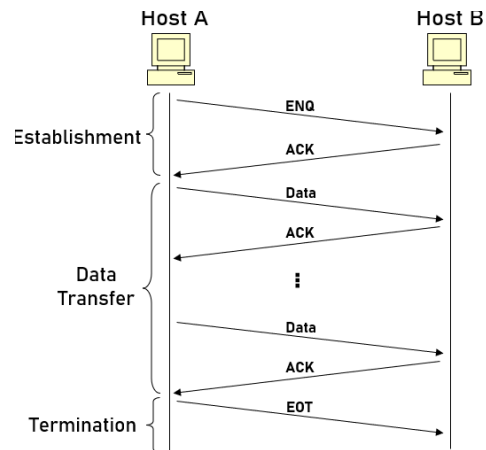


Figure 6.10 Line Discipline (ENQ/ACK)

For connection of establishment, Host B responds either with ACK or NAK. The Host A tries to send ENQ three times before concluding that Host B is down.

Line Discipline (Poll / Select)

In case of multipoint connections there is one primary and multiple secondary hosts. It's important to note that communication between secondary devices goes over the primary. **Select** mode is used when primary has something to send to a secondary (downstream). The **Poll** mode is used to solicit transmissions from a secondary to the primary (upstream). Address must be contained in all packets.

In the Select mode, the SEL packet contains address of B. B can respond either by ACK or NAK. The primary sends one or more data packets, which are acknowledged (ACK) by B.

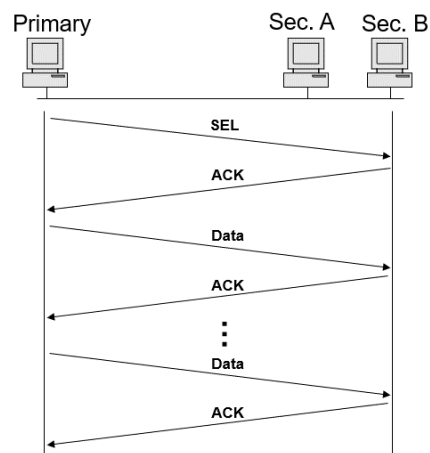


Figure 6.11 Line Discipline (Poll/ Select)

Poll mode

In the Poll packet contains address of the recipient. If the intended secondary has no data to send, replies with NAK. The Data is ACKed by the primary. It is important here to note how to terminate the connection?

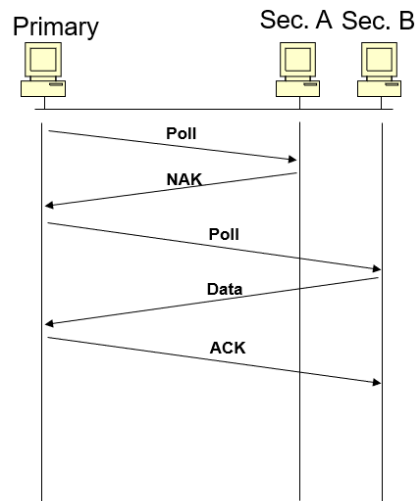


Figure 6. 12 Line Discipline (Poll / Select) with NAK

Flow Control

It restrict the amount of data to send before waiting for acknowledgment categories:

- i) **Stop-and-Wait** Wait for an ACK before sending the next frame
- ii) **Continuous (Sliding Window)** Can send several windows before requiring an ACK

i) Stop-and-Wait Flow Control Technique

In this approach a sender sends a frame, and then wait until you get ACK back. In a noisy communication channel, if a frame is damaged in transit, the receiver hardware will detect this when it computes the checksum. If a damaged frame is received, it will be discarded, and transmitter will retransmit the same frame after receiving a proper acknowledgement. If the acknowledgement frame gets lost and data link layer on 'A' eventually times out. Not having received an ACK, it assumes that its data frame was lost or damaged and sends the frame containing packet 1 again. This duplicate frame also arrives at data link layer on 'B', thus part of file will be duplicated and protocol is said to be failed.

A typical approach to solve this problem is the provision of a sequence number in the header of the message. The receiver can then check the sequence number determine if the message is a duplicate since only message is transmitted at any time. The sending and receiving station needs only 1-bit alternating sequence of '0' or '1' to maintain the relationship of the transmitted message and its ACK/NAK. A modulo-2 numbering scheme is used where the frames are alternatively label with '0' or '1' and positive acknowledgements are of the form ACK 0 and ACK 1.



Sequence Numbers

The protocol specifies that frames need to be numbered. This is done by using sequence number. A field is added to the data frame to hold the sequence number of that frame. The sequence numbers are based on modulo-2 arithmetic. Stop-and-wait ARQ is the simplest mechanism for error and flow control.

Advantage:

The main advantage of this approach is it's Simplicity

Disadvantage:

The main disadvantage of this approach is it's inefficiency



Given a **1-bit frame**, a **b bits/sec** transmission speed, a **R seconds** roundtrip time, and an **infinitely short ACK**. The value of utilization can be calculated as $Utilization = 1 / (1 + bR)$. It should be noted that it is not suited for high bandwidth-delay product lines.

ii) Continuous (Sliding Window) Flow Control Technique

It sends multiple frames before waiting for ACK. It is important to note that several frames can be in transit at a time. There are two types of windows, the sender window vs. receiver window

- Frames can be ACKed without waiting for the receiver window to fill up
- Frames may be sent as long as there are unsent packets in the sender window
- Sliding window scheme uses the modulo-n arithmetic
- The addresses range between 0 and n-1

The size of the window is n-1 to prevent ambiguity in ACKs, which contains sequence number of next expected frame.

- Let us assume that the window size is n-1
- Sender can send frames 0 to n-2 to start
- Receiver ACKs frame 0
- Does it mean that it is ready for the next frames?

Sender window:

- Shrink from left as frames are sent
- Expand from right as you receive ACKs

Receiver window:

- Shrink from left as frames are received
- Expand from right as ACKs are sent

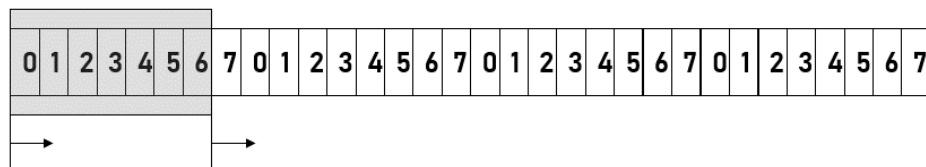


Figure 6. 13 Shrinking and Expanding of Sender and Receiver window in Sliding Window Protocol

Flow Control (Error Control)

Error control in data link layer is based on retransmissions (backward Error Control). Any time an error is detected, a NAK is returned to sender. It can be categorized into three categories:

a) Automatic Repeat Request (ARQ)

- Retransmission in case of damaged frame, lost frame, and lost ACK

b) Stop-and-Wait ARQ

c) Sliding Window ARQ

- Go-Back-n
- Selective Repeat

a) Automatic Repeat Request (ARQ)

Automatic Repeat Request (ARQ) is an error-control mechanism for data transmission which uses acknowledgements (or negative acknowledgements) and timeouts to achieve reliable data transmission over an unreliable communication link.

b) Flow Control (Stop-and-Wait ARQ)

In this approach, the sender window size is 1. This allows the sender to keep only one frame unacknowledged. So, sender sends one frame and then **waits** until the sent frame gets acknowledged. After receiving the acknowledgement from the receiver, sender sends the next frame. Number frames as 0 and 1 alternately. Keep a copy of the last sent frame. The receiver

acknowledges the receipt of frame 0 by sending ACK 1 (waiting for frame 1). If NAK is received, resend the last frame (no numbering necessary). It takes care of damaged data frames. It starts a timer for every frame sent. If it expires before an ACK is received, resend the frame. It takes care of lost data frame and lost ACKs / NAKs. The receiver discards duplicate frames in case ACK is lost

c) Flow Control (Sliding Window ARQ)

The sliding window protocols are the data link layer protocols for reliable and sequential delivery of data frames. The sliding window is also used in Transmission Control Protocol.

In this protocol, multiple frames can be sent by a sender at a time before receiving an acknowledgment from the receiver. The term sliding window refers to the imaginary boxes to hold frames. Sliding window method is also known as windowing.

How it Works?

In these protocols, the sender has a buffer called the sending window and the receiver has buffer called the receiving window.

The size of the sending window determines the sequence number of the outbound frames. If the sequence number of the frames is an n -bit field, then the range of sequence numbers that can be assigned is 0 to $2^n - 1$. Consequently, the size of the sending window is $2^n - 1$. Thus in order to accommodate a sending window size of $2^n - 1$, a n -bit sequence number is chosen.

The sequence numbers are numbered as modulo- n . It keeps a copy of all unACKed frames. It receives single ACK for multiple frames.



If last ACK was for 2 (waiting for 3) and new ACK is for 6, frames 3, 4, 5, and 6 are ACKed.

Timers are used to signal retransmissions.

The Response to NAKs and expiring timers determine the style of ARQ

i) *Go-Back- n*

ii) *Selective Repeat*

i) Flow Control (Go-Back- n ARQ)



Let us see how it deals with the damaged frames. Let us assume that frames 0-5 are sent and all but frame # 3 is correctly received. The receiver sends NAK 3 and discards subsequent frames. This signals to sender that frames 0, 1, and 2 are correctly received and frame 3 is damaged. The sender retransmits frames 3, 4, and 5. The same procedure is followed for lost data frames.

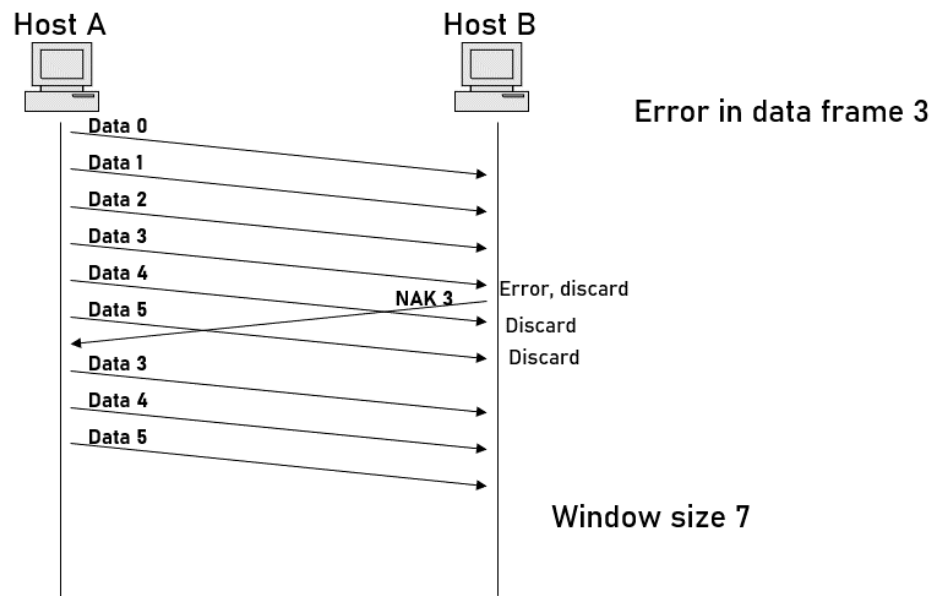


Figure 6. 14 Lost or damaged frame



Let us now understand, how it deals with the lost Acknowledgements. When sender reaches window capacity, it starts a timer. If timer expires, it resends all outstanding (unACKed) frames. The receiver discards possible duplicate frames and sends another ACK

Lost ACK:

In case of a lost acknowledgement, the sender will assume that the frame is lost and will retransmit the frame.

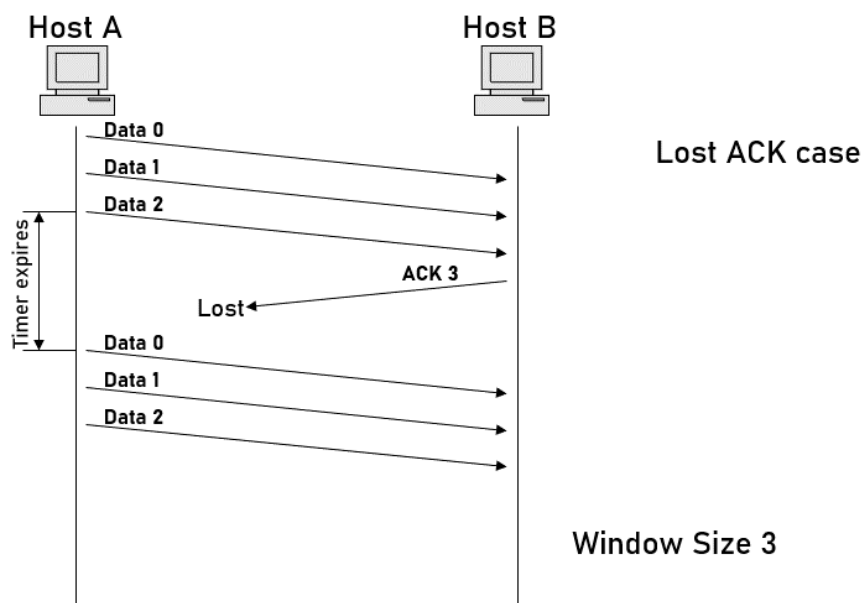


Figure 6. 15 Lost or damaged acknowledgement

ii) Selective Repeat ARQ

It should be noted that the go-back-n protocol works well if errors are less, but if the line is poor it wastes a lot of bandwidth on retransmitted frames. An alternative strategy, the selective repeat protocol, is to allow the receiver to accept and buffer the frames following a damaged or lost one. Selective Repeat attempts to retransmit only those packets that are actually lost (due to errors).

Receiver must be able to accept packets out of order. Since receiver must release packets to higher layer in order, the receiver must be able to buffer some packets.

Data Link Control Protocols - Noisy Channels (Go-Back-N ARQ Protocol)

Pipelining

It is very useful that a task is begun before the previous task has ended. This is known as pipelining. There is no pipelining in Stop-and-Wait ARQ. Pipelining does apply to other two Noisy channel protocols. Here several frames can be sent before we receive news about the previous frames. Pipelining improves the efficiency of the transmission. To improve the efficiency of transmission, multiple frames must be in transition while waiting for acknowledgment. This is done to keep the channel busy while the sender is waiting for acknowledgment.

In Go-Back-N ARQ protocol, we can send several frames before receiving acknowledgments. We keep a copy of these frames until the acknowledgments arrive.

Sequence Numbers

The Frames from a sender station are numbered sequentially. As we need to include the sequence number of each frame in the header, we need to set a limit. If the header of the frame allows m bits for the sequence number, the sequence numbers range from 0 to $2^m - 1$.



If m is 4, the only sequence numbers are 0 through 15 inclusive. We can repeat the sequence. So, the sequence numbers are:

0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, ...

In other words, the sequence numbers are modulo- 2^m .



Sliding Window - The sliding window defines the range of sequence numbers. The sender and receiver need to deal with only part of the possible sequence numbers:

- The range which is the concern of the sender is called the send sliding window.
- The range that is the concern of the receiver is called the receive sliding window.

Send Sliding Window

The send window is an imaginary box covering the sequence numbers of the data frames which can be in transit. In each window position, some of these sequence numbers define the frames that have been sent; others define those that can be sent. The maximum size of the window is $2^m - 1$. Window size may be fixed and set to the maximum value, but some protocols may have a variable window size.



Figure below shows a sliding window of size 15 ($m = 4$).

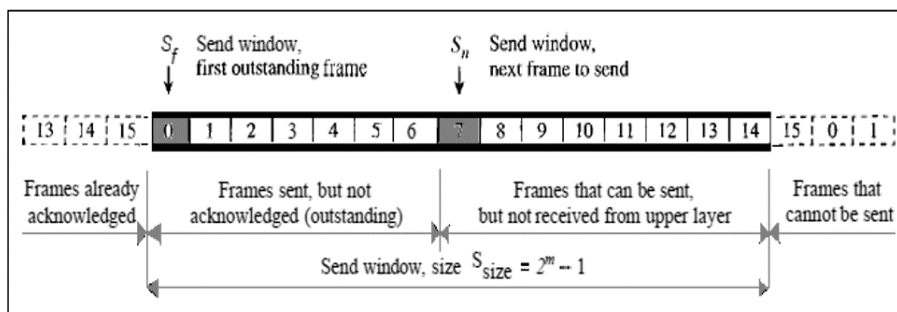


Figure 6. 16 Send window before sliding

Sliding Window

The window at any time divides the possible sequence numbers into four regions:

1. First region, from the far left to the left wall of the window, defines the sequence numbers belonging to frames already acknowledged.
 - Sender does not worry about these frames and keeps no copies of them.

2. Second region defines the range of sequence numbers belonging to the frames that are sent and have an unknown status.
 - Sender needs to wait to find out if these frames have been received or were lost. We call these outstanding frames.
3. Third range defines the range of sequence numbers for frames that can be sent; however, the corresponding data packets have not yet been received from the network layer.
4. Fourth region defines sequence numbers that cannot be used until the window slides.

Three variables define window size and location at any time:

- **Sf** (send window, the first outstanding frame)

Sequence number of the first (oldest) outstanding frame.

- **Sn** (send window, the next frame to be sent)

Sequence number that will be assigned to the next frame to be sent.

Ssize (send window, size) – It defines the size of the window

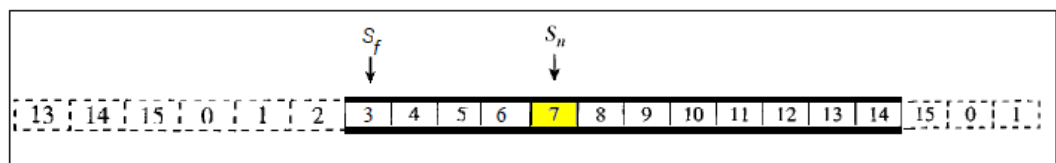


Figure 6.17 Send Window After Sliding

Figure above shows how a send window can slide one or more slots to the right when an acknowledgment arrives from the other end.

The acknowledgments in this protocol are cumulative, meaning that more than one frame can be acknowledged by an ACK frame.



The send window can slide one or more slots when a valid acknowledgment arrives.

In this figure, frames 0, 1, and 2 are acknowledged, so the window has slid to the right three slots.

Note that the value of Sf is 3 because frame 3 is now the first outstanding frame.

Receive Sliding Window

The receive window makes sure that the correct data frames are received and that the correct acknowledgments are sent. The size of the receive window is always 1. The receiver is always looking for the arrival of a specific frame. Any frame arriving out of order is discarded and needs to be resent. Figure below shows the receive window.

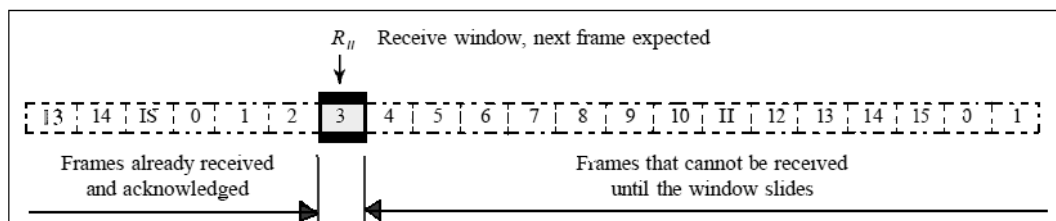


Figure 6.18 (a) The receive window.

Figure: (a)Receive Window

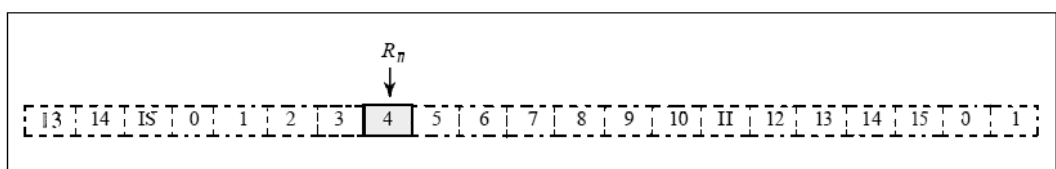


Figure 6. 19 (b) The receive window after sliding

Note that the receive window has one single variable R_n . The window slides when a correct frame has arrived; sliding occurs one slot at a time.

Receive Sliding Window

Timers

Although there can be a timer for each frame that is sent, in this protocol we use only one. The reason is that the timer for the first outstanding frame always expires first; we send all outstanding frames when this timer expires.

Acknowledgment

The receiver sends a positive acknowledgment if a frame has arrived safe and sound and in order. If a frame is damaged or is received out of order, receiver is silent and will discard all subsequent frames until it receives the one it is expecting. The silence of the receiver causes the timer of the unacknowledged frame at the sender site to expire. This causes the sender to go back and resend all frames, beginning with the one with the expired timer. The receiver does not have to acknowledge each frame received. It can send one cumulative acknowledgment for several frames.

Receive Sliding Window (Resending a Frame)

When the timer expires, the sender resends all outstanding frames.



Suppose the sender has already sent frame 6, but the timer for frame 3 expires. This means that frame 3 has not been acknowledged. The sender goes back and sends frames 3, 4, 5, and 6 again.

That is why the protocol is called Go-Back-N ARQ.

Sender Sliding Window

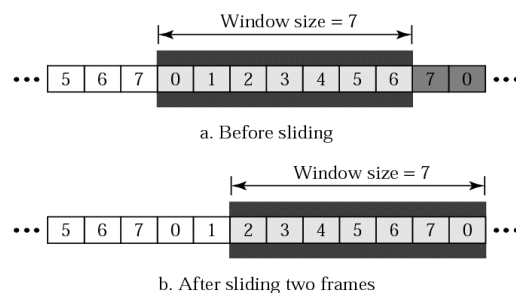


Figure 6. 20 Sender Sliding Window

Receiver Sliding Window

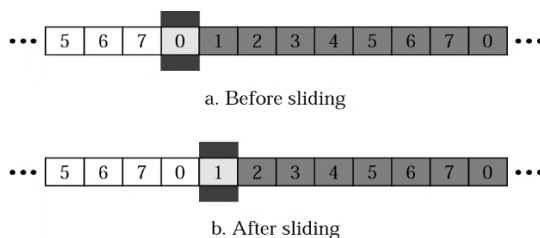


Figure 6. 21 Receiver Sliding Window

Normal Operation

The normal operation of the sliding window protocol can be seen in figure 6.22.

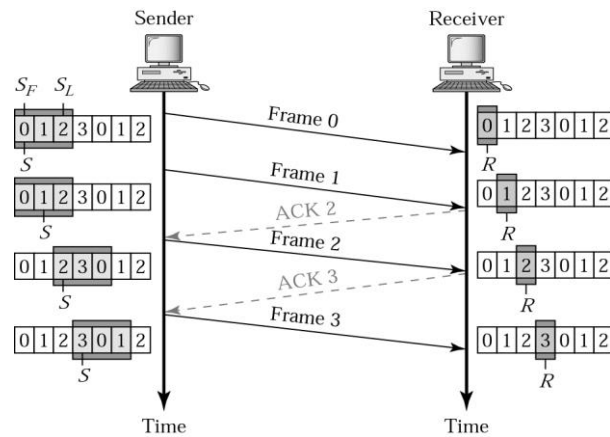


Figure 6.22 Normal Operation

Lost Frame

Let us see a scenario in which a frame gets lost. In Figure 6.23, it can be seen that the frame 2 is lost and no acknowledgement for it has been received. So after the time period, the frame is again sent to the receiver. You can see the same in figure 6.23

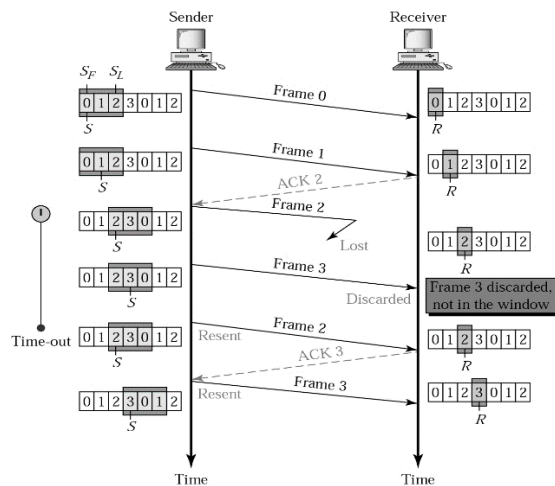


Figure 6.23 Lost Frame Scenario

Sender Window Size

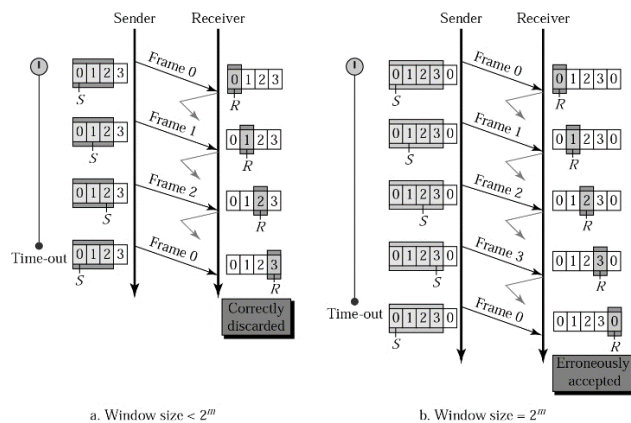


Figure 6.24 Sender Window Size



Size of the sender window must be $< 2^m$. Size of the receiver window is always 1.

Go-Back-N Sender Algorithm

```

21  if(Event (ArrivalNotification)) //ACK arrives
22  {
23      Receive(ACK);
24      if{corrupted{ACK}}
25          Sleep();
26      if ((ackNo>sf)&&(ackNo<=Sn)) //If a valid ACK
27          While(Sf <= ackNo)
28          {
29              PurgeFrame (Sf);
30              Sf = Sf + 1;
31          }
32          StopTimer();
33      }
34
35  if(Event(TimeOut)) //The timer expires
36  {
37      StartTimer();
38      Temp = Sf;
39      while(Temp < Sn);
40      {
41          SendFrame(Sf);
42          Sf = Sf + 1;
43      }
44  }
45  }

```

Figure 6. 25 Go-back-N Sender Algorithm

Go-Back-N Receiver Algorithm

```

1  Rn = 0;
2
3  while (true) //Repeat forever
4  {
5      WaitForEvent();
6
7      if(Event(ArrivalNotification)) //Data frame arrives
8      {
9          Receive(Frame);
10         if(corrupted(Frame))
11             Sleep( );
12         if(seqNo = Rn) //If expected frame
13         {
14             DeliverData(); //Deliver data
15             Rn = Rn + 1; //Slide window
16             SendACK(Rn);
17         }
18     }
19 }

```

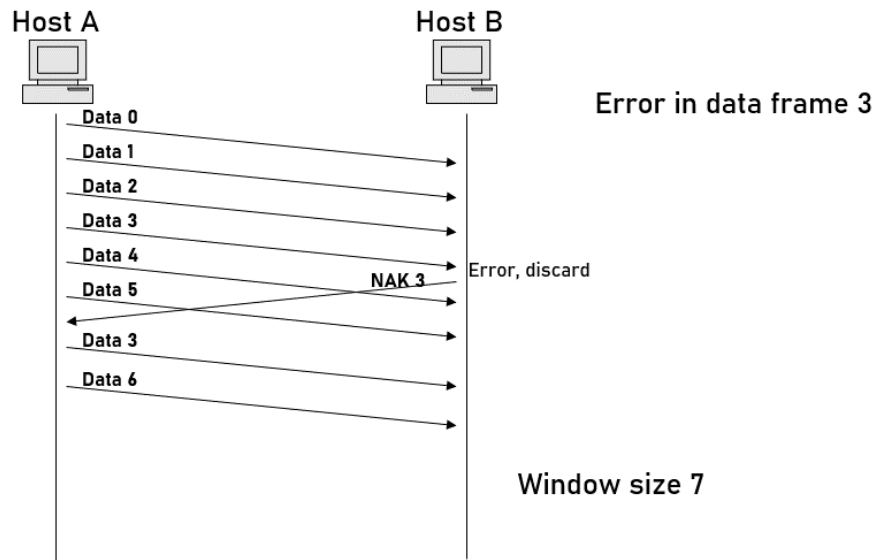
Figure 6. 26 Go-Back-N Receiver Algorithm

ii) Selective Repeat ARQ

Only damaged and lost frames are retransmitted. The ACK numbers refer to the last correctly received frame, not next frame expected. If go-back-n window size is n-1, SR widow size must be at most n/2.

Damaged Frames:

- After receiving a damaged frame, receiver sends a NAK and continues accepting other frames
- Sender only retransmits the missing frame



Lost Data Frames:

Here the out-of-sequence delivery is permitted, but out-of-sequence ACK is not permitted. So, when a lost frame is detected, NAK is sent. If last frame is lost, then receiver does nothing.

Lost ACK:

Here the lost acknowledgement is handled the same way as in go-back-n.

ii) Stop and Wait Automatic Repeat Request

The Stop-and-wait ARQ, is also referred to as alternating bit protocol. It is a method in telecommunications to send information between two connected devices. It ensures that information is not lost due to dropped packets and that packets are received in the correct order. It is the simplest automatic repeat-request mechanism. A stop-and-wait ARQ sender sends one frame at a time; it is a special case of the general sliding window protocol with transmit and receive window sizes equal to one in both cases. After sending each frame, the sender doesn't send any further frames until it receives an acknowledgement (ACK) signal. After receiving a valid frame, the receiver sends an ACK. If the ACK does not reach the sender before a certain time, known as the timeout, the sender sends the same frame again. The timeout countdown is reset after each frame transmission.

Keywords

Bluetooth: Bluetooth is a proprietary open wireless technology standard for exchanging data over short distances (using short wavelength radio transmissions in the ISM band from 2400-2480 MHz) from fixed and mobile devices, creating personal area networks (PANs) with high levels of security.

Go-Back-N ARQ is a specific instance of the automatic repeat request (ARQ) protocol, in which the sending process continues to send a number of frames specified by a window size even without receiving an acknowledgement (ACK) packet from the receiver.

Sliding Window Protocol: is a feature of packet-based data transmission protocols. Sliding window protocols are used where reliable in-order delivery of packets is required, such as in the data link layer (OSI layer 2) as well as in the Transmission Control Protocol (TCP).

Protocol Verification: Protocols are verified either during the design phase before implementation of the system or during the testing and simulation phase after implementation of the system.

Selective Repeat: Provides buffers at source and destination hosts to enable the source node to have more than one outstanding frame at a time and destination node to accept out of order frames and store them in its window.

Simplex Stop and Wait: After transmission, the source node waits for an acknowledgement from the destination node. After receiving, the acknowledgement, the loop starts over again.

Self Assessment

Fill in the blanks:

- describes the techniques to access a shared communication channel and reliable transmission of data frame in computer communication environment.
- does not include any connection setup or release and does not deal with frame recovery due to channel noise.
- refers to a reliable transfer of bit streams to the network layer for which the data link layer breaks the bit stream into frames.
- controls mismatch between the source and destination hosts data sending and receiving speed and therefore dropping of packets at the receiver end.
- In stop and wait protocol, the acknowledgement frame has bits that the destination node sends back to the source machine.
- Positive Acknowledgement with Retransmission Protocol (PAR) uses to determine if any frames is lost or damaged.
- The Go Back N protocol overcomes the problem of PAR by enabling the source machine to have more than at a time by using buffers.

Multiple Choice Questions:

- Stop-and-wait ARQ is a _____ technique.
 - line discipline
 - Error control
 - flow control
 - Session management
- Which of the following is true regarding the Send Sliding Window?
 - Send window is an imaginary box covering the sequence numbers of the data frames which can be in transit.
 - In each window position, some of these sequence numbers define the frames that have been sent; others define those that can be sent.
 - The maximum size of the window is 2m.
 - Window size may be fixed and set to the maximum value, but some protocols may have a variable window size.
- The sender which sends one frame and then waits for an acknowledgement before process is known as _____?
 - Stop and Wait ARQ.
 - Simplest protocol
 - Selective Repeat ARQ
 - Go-back-N ARQ
- In the Stop-and-Wait ARQ, if the acknowledgement frame gets lost, then not having received an ACK, it assumes that _____.
 - Data frame is lost or damaged
 - ACK is lost
 - Both Data frame and ACK are lost or damaged
 - None of the given choices
- In Stop and Wait Automatic Repeat Request, the arrival of the duplicate message can be identified with _____.
 - Adding a sequence number in the header of the message.
 - Adding a sequence number in the trailer of the message.
 - Checking the ACK of the message
 - None of the given choices

13. In Stop and Wait Automatic Repeat Request, a NAK denotes that the _____
 - a) data packet is found to be corrupt.
 - b) data packet has arrived out of sequence.
 - c) data packet has arrived late
 - d) None of the given choices
14. The Stop and Wait Automatic Repeat Request, is the simplest mechanism for _____
 - a) error control
 - b) flow control
 - c) error control and flow control
 - d) None of the given choices
15. Noiseless channel is an ideal channel in which no frames are _____
 - a) lost
 - b) duplicated
 - c) corrupted
 - d) lost, duplicated, or corrupted.

Review Questions

1. What is the data link protocol?
1. What advantages does Selective Repeat sliding window protocol offer over Go Back N protocol?
2. What is the purpose of flow control?
3. Describe how does finite state machine model carry out protocol verification.
4. What are different data link protocols available? Why does PPP have become popular?
5. How does the data link layer accomplish the transmission of data from the source network layer to the destination network layer?
6. What are the major advantages of the Stop and Wait Automatic Repeat Request.
7. Explain how damaged frames are managed in the Selective Repeat ARQ.

Answers: Self Assessment

- | | |
|--------------------------|--|
| 1. Data link layer | 2. Unacknowledged connectionless service |
| 3. Framing | 4. Rate of data transmission |
| 5. nil | 6. a sequence number |
| 7. one outstanding frame | 8. B |
| 9. C | 10. A |
| 11. A | 12. A |
| 13. A | 13. C |
| 15. D | |

Further Readings



Andrew S. Tanenbaum, *Computer Networks*, Prentice Hall.

Behrouz A. Forouzan and Sophia Chung Fegan, *Data Communications and Networking*, McGraw-Hill Companies.

J. D. Spragins, *Telecommunications Protocols and Design*, Addison-Wesley.

Rajneesh Agrawal and Bharat Bhushan Tiwari, *Computer Networks and Internet*, Vikas Publication



<https://www.geeksforgeeks.org/basics-computer-networking/>

Unit 07: Data Link Layer - Medium Access Control

CONTENTS

Objectives

After this lecture, you would be able to

Introduction

7.1 Sublayers of Data Link Layer

7.2 Multiple Access Control

7.3 Random Access

7.4 Carrier Sense Multiple Access with Collision Detection (CSMA/CD)

Summary

Keywords

Self Assessment

Objectives

After this lecture, you would be able to

- know the various Multiple Access Protocols.
- understand the Pure Aloha Protocol and how it is different from the Slotted Aloha Protocol.
- understand the concept of collision.
- understand the CSMA protocol and know its types
- understand the behavior of three persistent methods

Introduction

The Data Link Layer (DLL) acts as an interface between in the network layer and the physical layer of the OSI model. The data link layer is for the subdivided into two sub layers of protocols. These are the medium access control (MAC) and the Logical Link Control (LLC). The MAC sublayer is responsible for some critical roles during data transmission. It's main task is to allocate Medium Access to the contending Nodes. The Logical Link Control (LLC) is towards the top of the MAC layer and performs some very critical functions like the cyclic redundancy check (CRC).

7.1 Sublayers of Data Link Layer

The Data link layer divided into two functionality-oriented sublayers as shown in Figure 7.1. These sublayers are:

- The upper sublayer is responsible for datalink control,
- The lower sublayer is responsible for resolving access to the shared media.

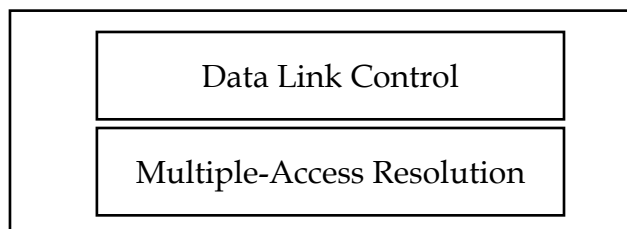


Figure 7. 1 Sublayers of Data Link layer

7.2 Multiple Access Control

These protocols are called Medium or Multiple Access Control (MAC) Protocols belong to a sublayer of the data link layer called MAC (Medium Access Control). When nodes or stations are connected and use a common link, called a multipoint or broadcast link. So, we need a multiple-access protocol to coordinate access to the link. Figure 7.2 Shows the different types of Multiple Access Protocols.

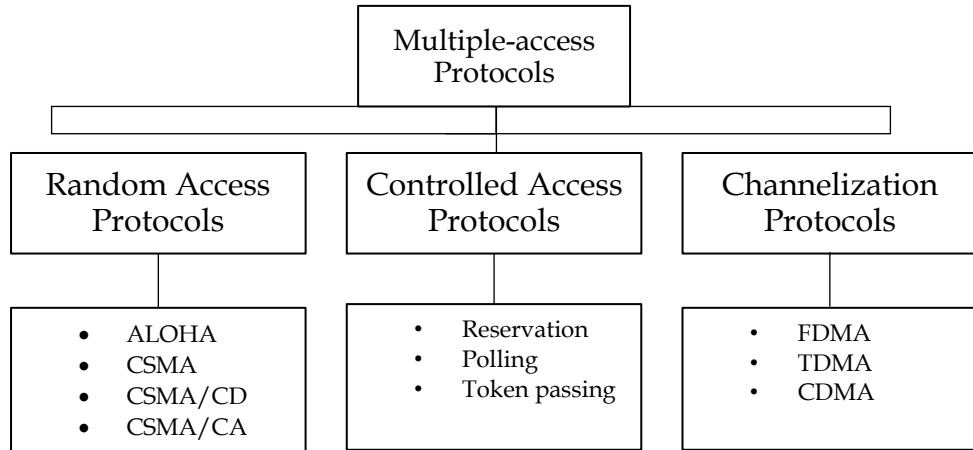


Figure 7. 2 Multiple Access Protocols

In random access or contention methods, no station is superior to another station and none is assigned the control over another. No station permits, or does not permit, another station to send. At each instance, a station that has data to send uses a procedure defined by the protocol to make a decision on whether or not to send.

7.3 Random Access

If more than one station tries to send, there is an access conflict i.e. COLLISION, the frames will be either destroyed or modified.

To avoid access conflict, each station follows a procedure.

- When can the station access the medium?
- What can the station do if the medium is busy?
- How can the station determine the success or failure of the transmission?
- What can the station do if there is an access conflict?

Random Access Control Protocols

In these protocols, all stations have same superiority that is no station has more priority than another station. Any station can send data depending on medium's state (idle or busy). Also, there is no fixed sequence of stations sending data. It has the following types.

a) **ALOHA**

b) **Carrier Sense Multiple Access (CSMA)**

- i) Carrier Sense Multiple Access with Collision Detection (CSMA/CD)
- ii) Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA)

ALOHA

A very interesting protocol known as ALOHA, which used a very simple procedure called multiple access (MA), the earliest random-access method, was developed at the Univ. of Hawaii in the early 1970s.

- Base station is central controller
- Base station acts as a hop

- Potential collisions, all incoming data is @ 407 MHz

ALOHA is a multiple access protocol for transmission of data via a shared network channel. It operates in the medium access control sublayer (MAC sublayer) of the open systems interconnection (OSI) model. Using this protocol, several data streams originating from multiple nodes are transferred through a multi-point transmission channel.

In ALOHA, each node or station transmits a frame without trying to detect whether the transmission channel is idle or busy. If the channel is idle, then the frames will be successfully transmitted. If two frames attempt to occupy the channel simultaneously, collision of frames will occur and the frames will be discarded. These stations may choose to retransmit the corrupted frames repeatedly until successful transmission occurs.

The Figure 7.3 depicts the working of a typical ALOHA protocol where all nodes are sending data at the same time

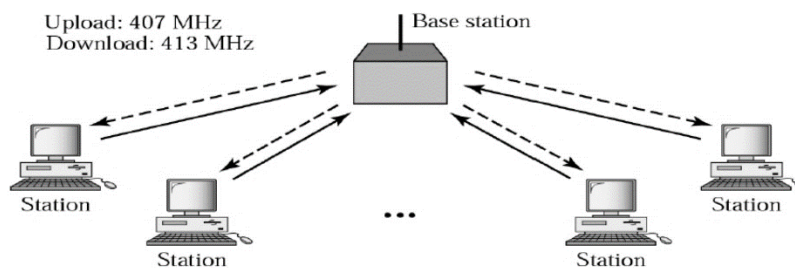


Figure 7.3 Working of Aloha Protocol

The original ALOHA protocol is a simple, but elegant protocol. The idea is that each station sends a frame whenever it has a frame to send. Figure 7.3 shows an example of frame collisions. ALOHA was designed for a radio (wireless) LAN, but it can be used on any shared medium. It is obvious that there are potential collisions in this arrangement. The medium is shared between the stations. When a station sends data, another station may attempt to do so at the same time. The data from the two stations collide and become garbled.

The ALOHA is categorized into two types of protocols:

- i) *Pure ALOHA Protocol*
- ii) *Slotted ALOHA Protocol*

i) *Pure ALOHA Protocol*

In Pure ALOHA. Some of these frames collide because multiple frames are competing or contending for the shared channel.

1. Each station sends a frame whenever is has a frame to send
2. One channel to share, possibility of collision between frames from different stations

When a station sends data, it waits for an acknowledgement. If the acknowledgement doesn't come within the allotted time, then the station waits for a random amount of time called back-off time (T_b) and re sends the data. Since different stations wait for different amount of time the probability of future collision decreases. The throughput of Pure Aloha is maximized when frames are of uniform length. Whenever the frames try to occupy the channel at the same time, there will be a collision, and both will be garbled. If the first bit of a new frame overlaps with just the last bit of a frame almost finished, both frames will be totally destroyed, and both will have to be retransmitted later.

$$\text{Vulnerable Time} = 2 \times T$$

Throughput = $G \times e^{-2G}$; where G is the number of stations wishing to transmit at the same time. Now let us see a scenario. As you can see in Figure 7.4, there are 4 stations transmitting the frames. Frame 1 transmits the frame without any collisions as no other station is transmitting frames at the same time.

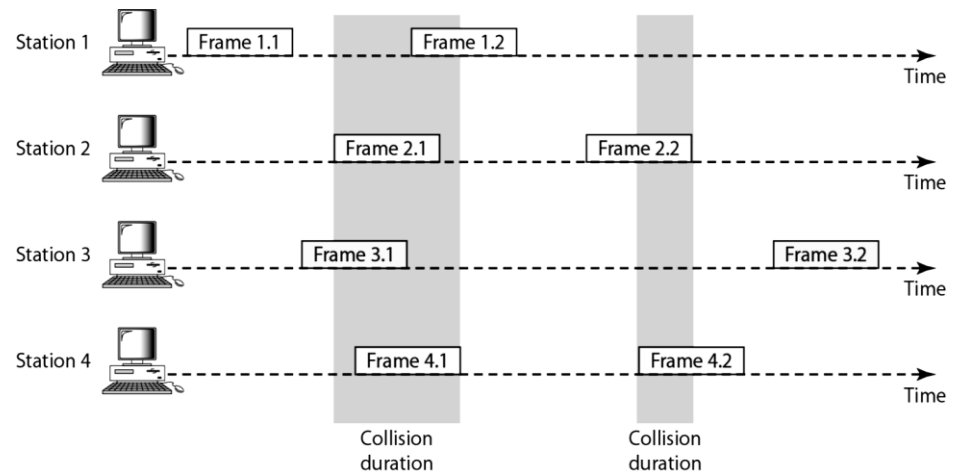


Figure 7.4 A scenario of Collisions

Now let us have a look at the first collision duration. After sending the frame, the station waits for an acknowledgment. If it does not receive an acknowledgment during the 2 times of the maximum propagation delay between the two widely separated stations ($2 \times T_P$), it assumes that the frame is lost; it tries sending again after a random amount of time ($T_B = R \times T_P$, or $R \times T_{fr}$).

The random time is calculated with the formulae:

$$T_B = R \times T_P \text{ or } R \times T_{fr}$$

- T_P (Maximum propagation time) = distance / propagation speed
- T_B (Back off time) : common formula is binary exponential back-off
- R is random number chosen between 0 to $2^k - 1$
- K is the number of attempted unsuccessful transmissions
- T_{fr} (the average time required to send out a frame)

Procedure for pure ALOHA protocol

The flow chart shown in Figure 7.5 shows the procedure of the Pure ALOHA protocol.

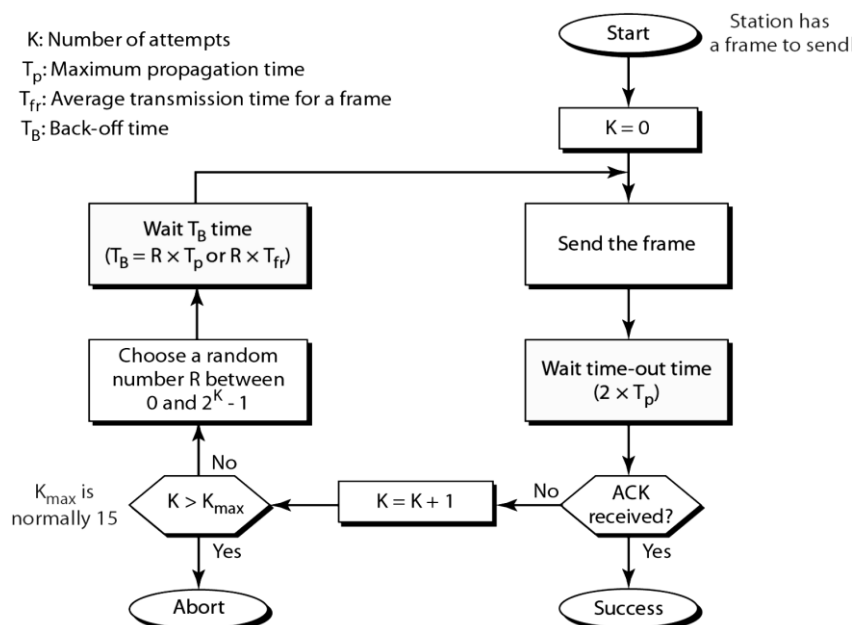


Figure 7.5 Procedure for Pure ALOHA Protocol

Vulnerable Time for Pure ALOHA Protocol

The vulnerable time of the Pure ALOHA Protocol is the time period during which collisions can occur. It can be understood by the depiction in Figure 7.6.

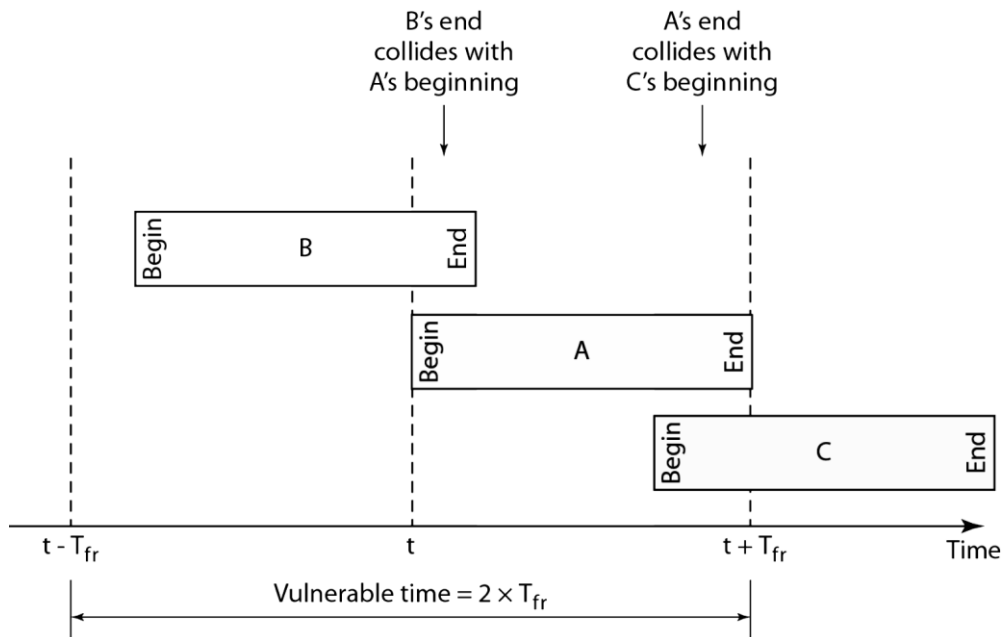


Figure 7.6 Vulnerable Time for Pure ALOHA Protocol

ii) Slotted ALOHA Protocol

In slotted ALOHA the time is divided into slots of T_{fr} and force the station to send only at the beginning of the time slot. This can be seen in Figure 7.7

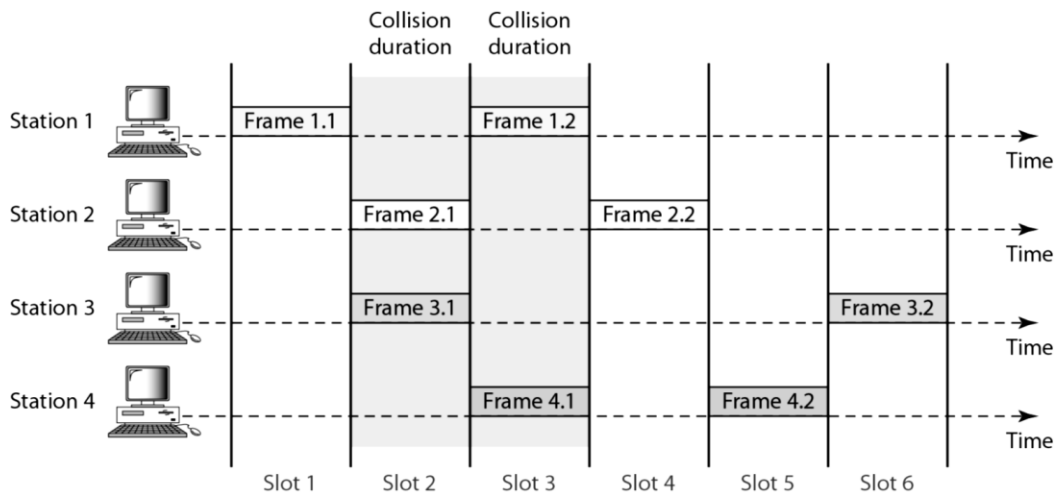


Figure 7.7 Slotted ALOHA Protocol

Vulnerable Time for Slotted ALOHA Protocol

The vulnerable time for slotted ALOHA protocol is vulnerable time = T_{fr} . The collision time periods can be clearly understood from Figure 7.8.

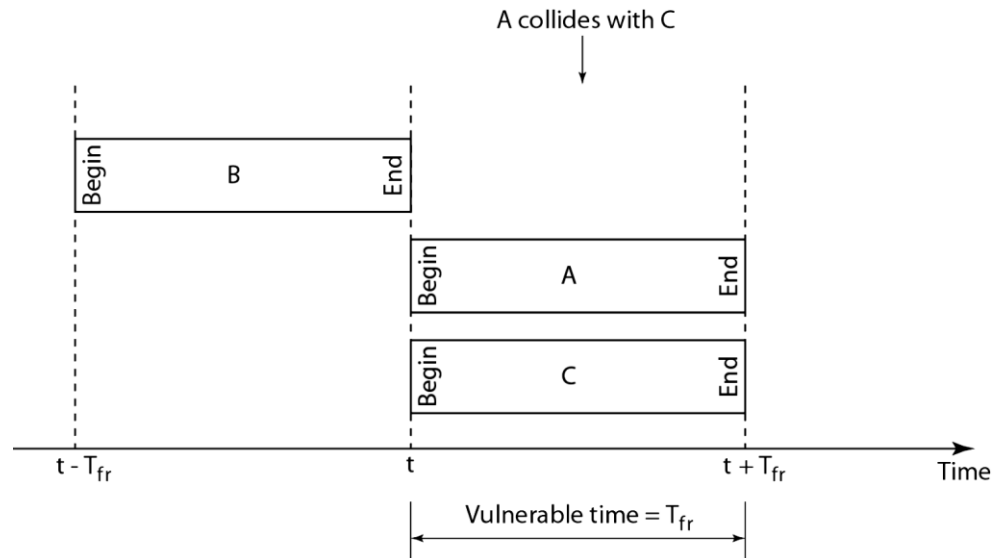


Figure 7. 8 Slotted ALOHA vulnerable time = T_{fr}

Pure ALOHA vs Slotted ALOHA

In case of the Pure Aloha the stations can transmit the available data at arbitrary time and colliding frames if any are destroyed. However in case of slotted Aloha a station has to wait for the beginning of the next time slot to start transmitting the important feature of slotted Aloha is that in this case the one rebel time period is decreased to almost half as compared to pure Aloha. Let us explore the differences between the two protocols with the help of Table 7.1.

Table 7. 1 Pure vs Slotted ALOHA Protocol

Pure ALOHA	Slotted ALOHA
Any station can transmit the data at any time.	Any station can transmit the data at the beginning of any time slot
The time is continuous and is not globally synchronised.	The time is discrete and globally synchronized
Vulnerable time in which collision may occur $= 2 \times T_{fr}$	The vulnerable time in which collisions may occur $= T_{fr}$
Probability of successful transmission of data packet $= G \times e^{-2G}$	Probability of successful transmission of data packet $= G \times e^{-G}$
Maximum efficiency 18.4% (Occurs at $G = 1/2$)	Maximum efficiency=36.8%(occurs at $G=1$)
Main advantage is the simplicity in implementation	Main advantage is that it reduces the number of collisions to half and doubles the efficiency of Pure Aloha

Carrier Sense Multiple Access (CSMA)

To minimize the chance of collision and, therefore, increase the performance, the CSMA method was developed. The chance of collision can be reduced if a station senses the medium before trying to use it. CSMA requires that each station first listen to the medium (or check the state of the medium) before sending. it is based on the principle "sense before transmit".

CSMA can reduce the possibility of collision, but it cannot eliminate it. The possibility of collision still exists because of propagation delay; a station may sense the medium and find it idle, only because the first bit sent by another station has not yet been received.

Space/Time model of the collision in CSMA

CSMA has the capability of reducing the possibility of collision. However it doesnot have the capability to eliminate it. The figure 7.9 shows a space and time model of a CSMA network

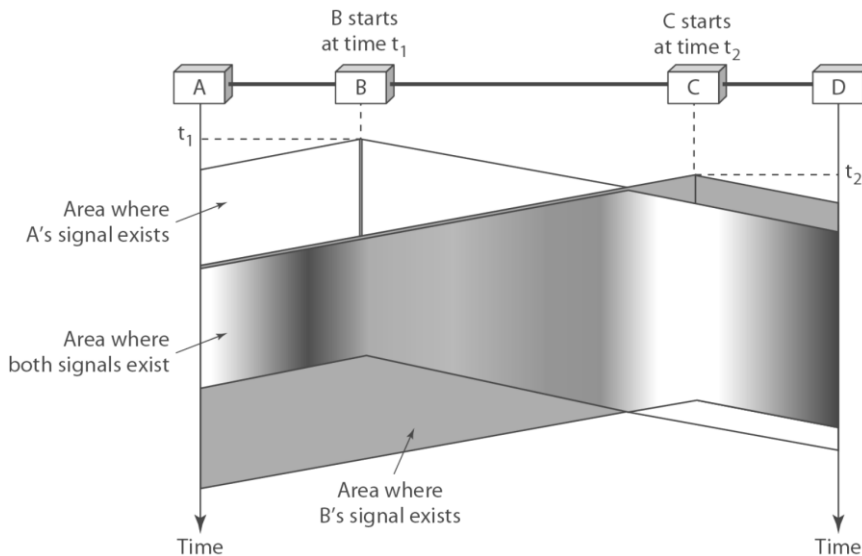


Figure 7. 9 A Space and Time model of a CSMA network

Vulnerable Time in CSMA

The vulnerable time in CSMA is the time period during which collisions can occur in the network. It can be clearly seen in Figure 7.10

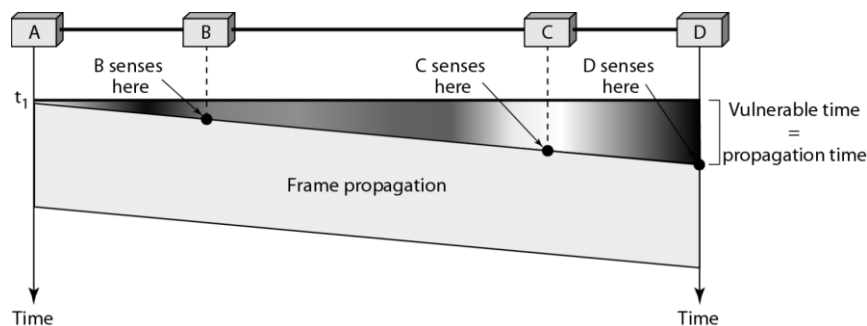


Figure 7. 10 Vulnerable Time in CSMA

Types of CSMA methods

Different CSMA methods that determine:

- What a station should do when the medium is idle?
- What a station should do when the medium is busy?

Three persistence strategies have been devised to answer these questions:

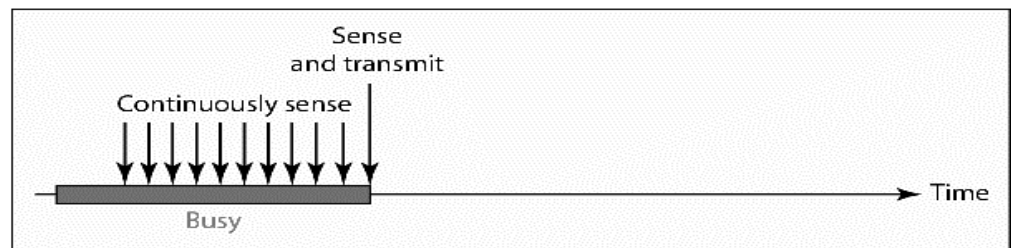
- i) **1-persistent method:** after the station finds the line idle, it sends its frame immediately (with probability 1)

- ii) **Non-persistent method:** a station that has a frame to send senses the line. If the line is idle, it sends immediately. If the line is not idle, it waits a random amount of time and then senses the line again.
- iii) **p-persistent method:** After the station finds the line idle, with probability p , the station sends its frames with probability $q=1-p$, the station waits for the beginning of the next time slot and checks the line again.

i) 1-Persistent Method

To avoid idle channel time, 1-persistent protocol used. Station wishing to transmit listens to the medium. If medium idle, transmit immediately; but if the medium is busy, continuously listen until medium becomes idle; then transmit immediately with probability 1.

Performance: In case of 1-Persistent method, the stations are selfish. If two or more stations becomes ready at the same time, collision guaranteed



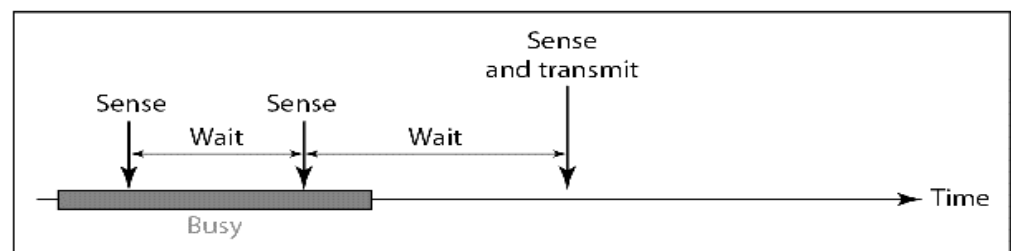
a. 1-persistent

ii) Non-Persistent Method

Here a station with frames to be sent, should sense the medium.

1. If medium is idle, **transmit**; otherwise, go to 2
2. If medium is busy, (**backoff**) wait a **random amount of time** and repeat 1
3. Non-persistent Stations are **deferential (respect others)**

Performance: In case of Non-Persistent method, random delays reduces probability of collisions because two stations with data to be transmitted will wait for different amount of times. Bandwidth is **wasted** if waiting time (backoff) is large because medium will remain idle following end of transmission even if one or more stations have frames to send



b. Nonpersistent

iii) p-Persistent Method

In this method, the time is divided to slots where each Time unit (slot) typically equals **maximum propagation delay**. For the station wishing to transmit listens to the medium;

If medium idle,

1. it will transmit with probability (p), OR
2. wait **one time unit (slot)** with probability ($1 - p$), then repeat 1.
3. If medium busy, **continuously listen until idle** and repeat step 1

Performance: It reduces the possibility of collisions like **nonpersistent**. It reduces channel idle time like **1-persistent**.

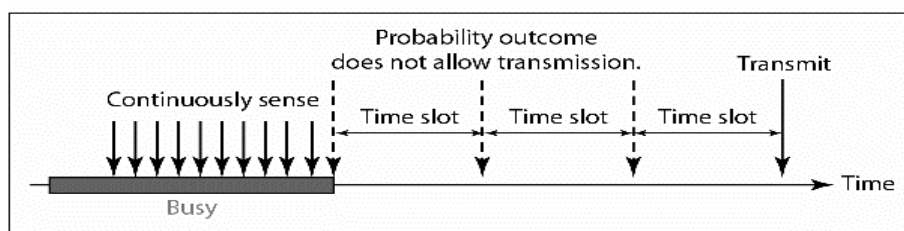


Figure 7.11 P-persistent

Flow Diagram for Three Persistence Methods

The flow diagram of the three persistence methods can be understood from figure 7.11

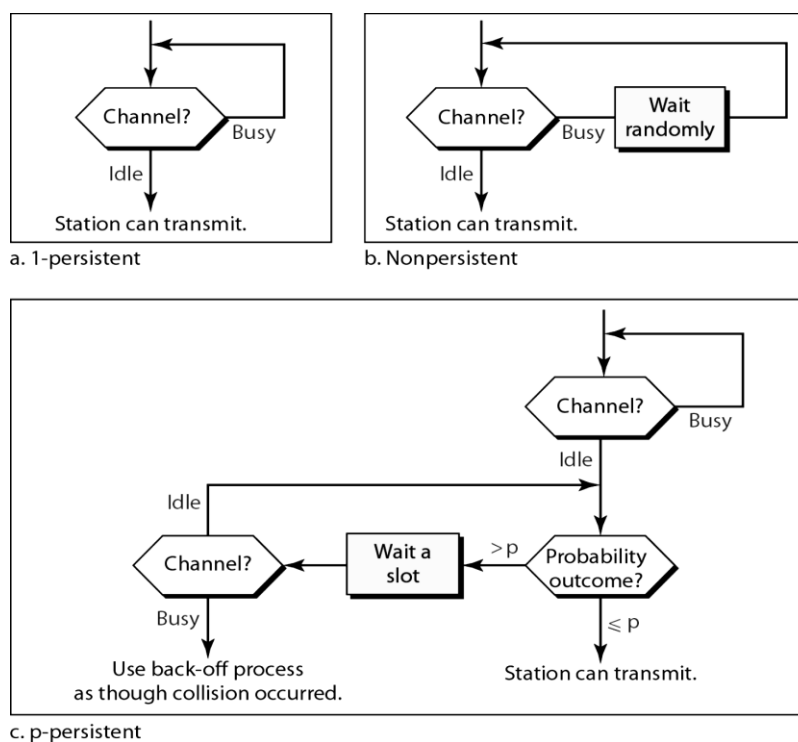


Figure 7.12 Flow Diagram for Three Persistence Methods

7.4 Carrier Sense Multiple Access with Collision Detection (CSMA/CD)

In this method, a station monitors the medium after it sends a frame to see if the transmission was successful. If so, the station is finished. If, however, there is a collision, the frame is sent again. In CSMA/CD, if 2 terminals begin sending packet at the same time, each will transmit its complete packet (although collision is taking place). Therefore, wasting medium for an entire packet time.

CSMA/CD performs the following steps

- Step 1: If the medium is idle, transmit
- Step 2: If the medium is busy, continue to listen until the channel is idle then transmit
- Step 3: If a collision is detected during transmission, cease transmitting
- Step 4: Wait a random amount of time and repeats the same algorithm

Figure 7.13 shows the collision of the first bit in CSMA/CD.

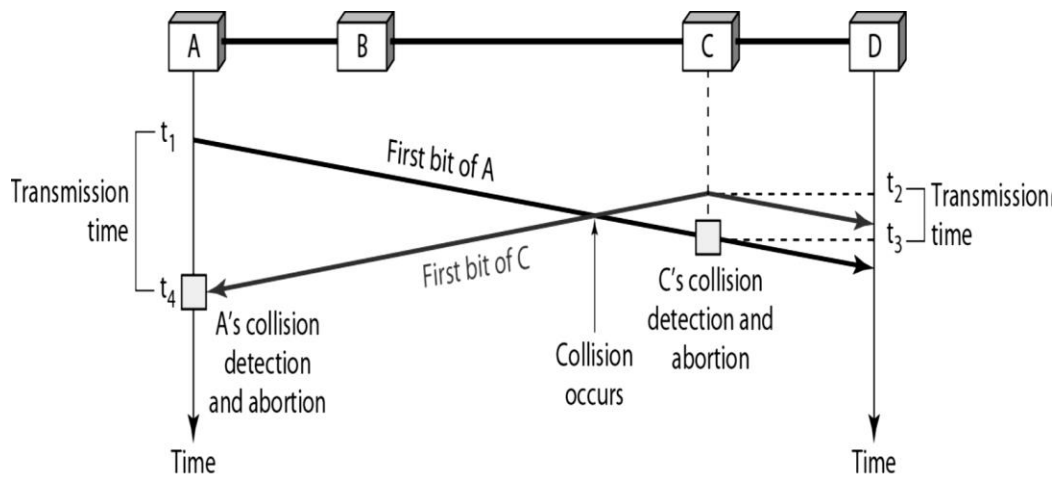


Figure 7.13 Performance: In case of Non-Persistent method

Collision and Abortion in CSMA/CD

The Figure 7.14 given below shows the collision and abortion in CSMA/CD.

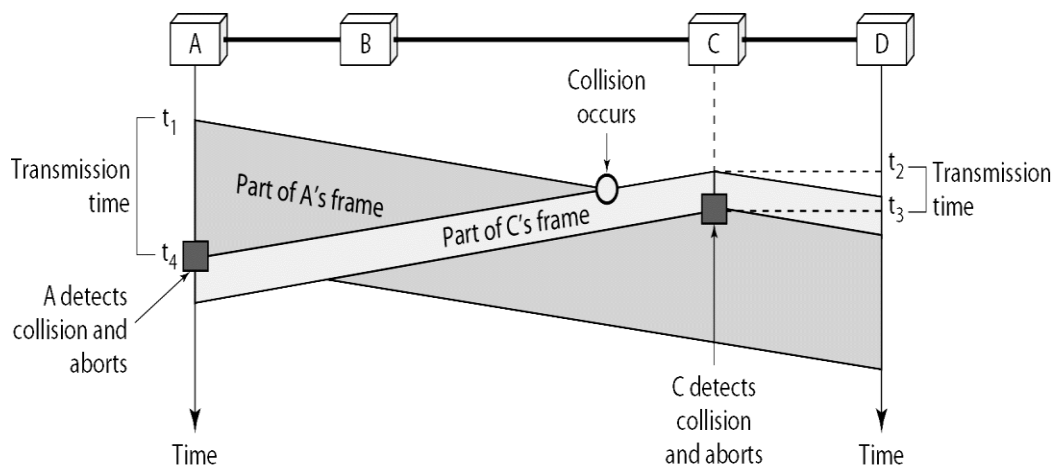


Figure 7.14 Collision and abortion in CSMA/CD

For CSMA/CD to work, we need a restriction on the frame size. The frame transmission time T_{fr} must be at least two times the maximum propagation time T_p .

Flow Diagram for the CSMA/CD

The figure 7.15 shows the flow diagram for CSMA/CD. It depicts the different actions taken in case of the different persistent methods.

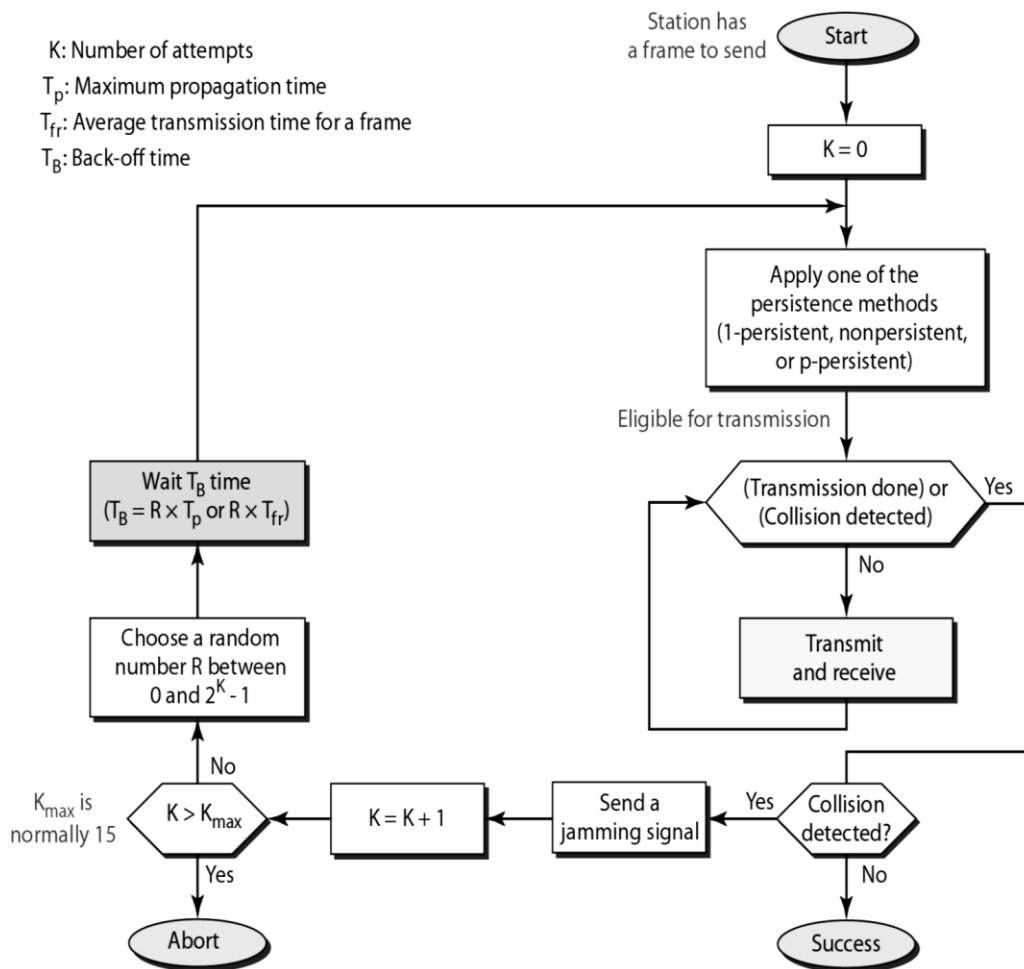


Figure 7.15 Flow Diagram for the CSMA/CD

Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA)

Collisions are avoided through the use of CSMA/CA's three strategies: the interframe space, the contention window, and acknowledgments.

- **Interframe Space:** Collisions are avoided by deferring transmission even if the channel is found idle. When an idle channel is found, the station does not send immediately. It waits for a period of time called the interframe space or IFS.
- **Contention Window:** The contention window is an amount of time divided into slots. A station that is ready to send chooses a random number of slots as its wait time. The number of slots in the window changes according to the binary exponential back-off strategy.
- **Acknowledgment:** With all these precautions, there still may be a collision resulting in destroyed data. In addition, the data may be corrupted during the transmission. The positive acknowledgment and the time-out timer can help guarantee that the receiver has received the frame.

All terminals listen to the same medium as CSMA/CD. The terminal ready to transmit senses the medium. If medium is busy it waits until the end of current transmission. It again waits for an additional predetermined time period DIFS (Distributed inter frame Space). when the counter reaches to zero. Then picks up a random number of slots (the initial value of backoff counter) within a contention window to wait before transmitting its frame. If there are transmissions by other terminals during this time period (backoff time), the terminal freezes its counter. It resumes count down after other terminals finish transmission + DIFS. The terminal can start its transmission when the counter reaches to zero. This can be seen in Figure 7.16.

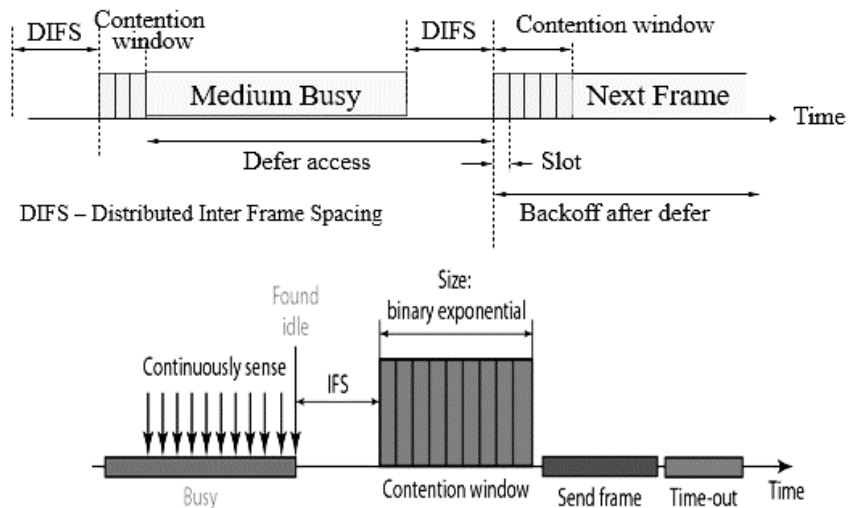


Figure 7. 16 Working of Carrier Sense Multiple Access with Collision Avoidance

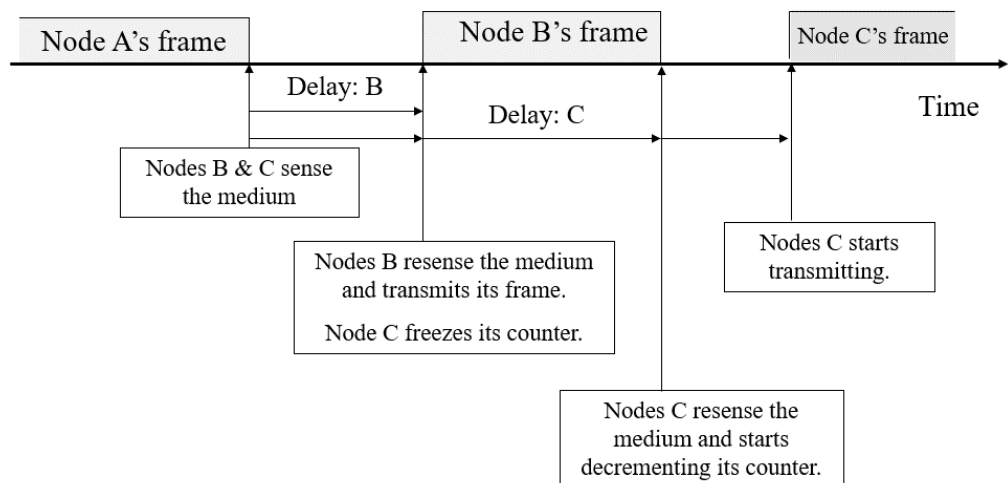


Figure 7. 17 Transmission of nodes in Carrier Sense Multiple Access with Collision Avoidance

The Carrier Sense Multiple Access with Collision Avoidance follows a series of Steps which can be very well seen in Figure 7.18.

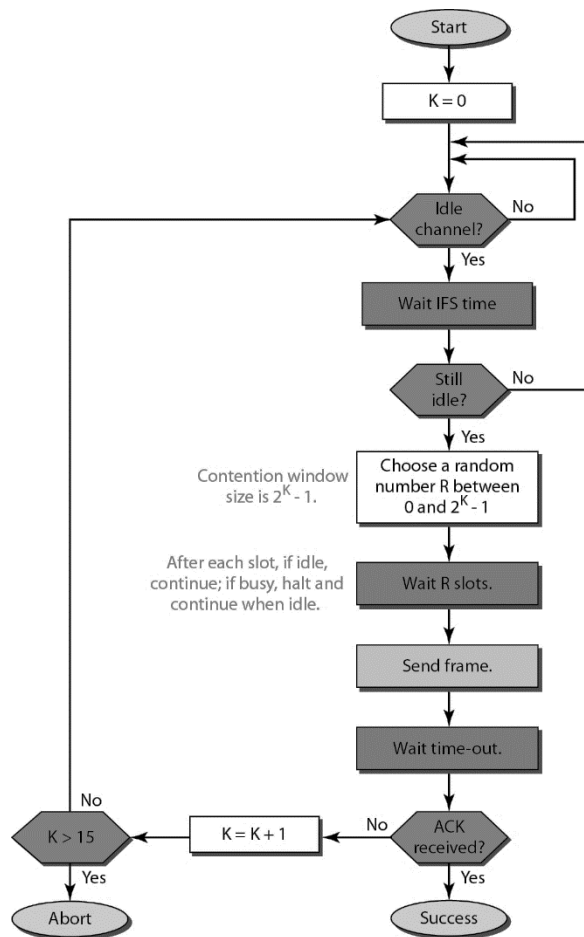


Figure 7. 18 Flow Chart of Carrier Sense Multiple Access with Collision Avoidance

CSMA/CA with ACK

The Carrier Sense Multiple Access with Collision Avoidance receives immediate acknowledgements from receiver upon reception of data frame without any need for sensing the medium. The ACK frame is transmitted after time interval SIFS (*Short Inter-Frame Space*) ($SIFS < DIFS$). It is important to note that the receiver transmits ACK without sensing the medium. If ACK is lost, retransmission is done. It is depicted in Figure 7.19

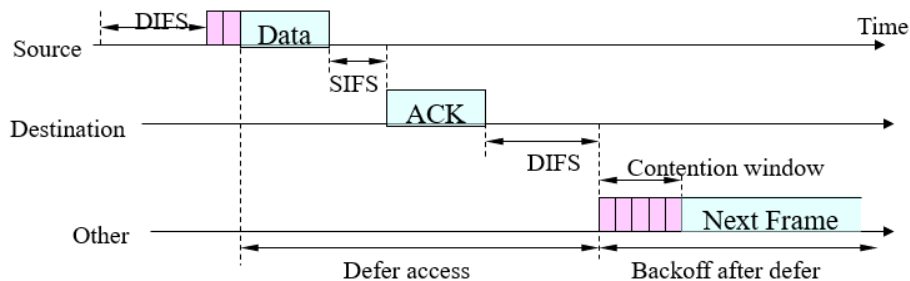


Figure 7. 19 CSMA/CA with ACK

CSMA/CA with RTS/CTS

The CSMA/CA with RTS/CTS Transmitter sends a RTS (request to send) after medium has been idle for time interval more than DIFS. The receiver responds with CTS (clear to send) after medium has been idle for SIFS. After this the data is exchanged. RTS/CTS is used for reserving channel for data transmission so that the collision can only occur in control message.

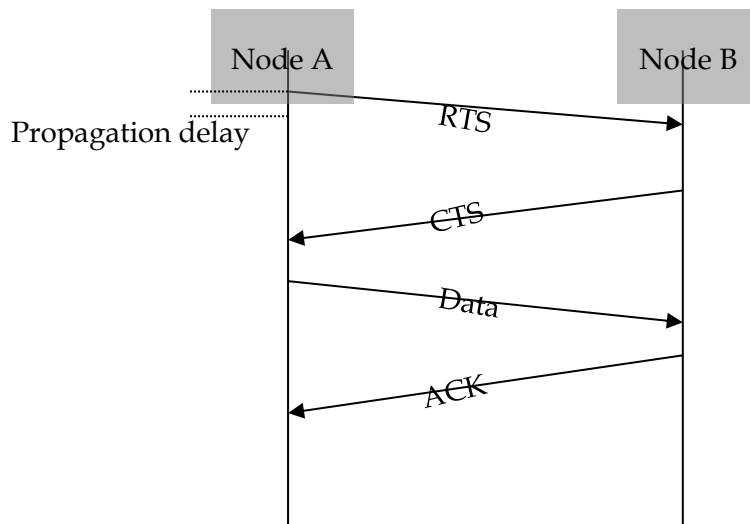


Figure 7. 19 CSMA/CA with RTS/CTS

Summary

The data link layer has the most important functions to perform like error control, flow control etc. It acts as an interface between the network layer and the physical layer of the OSI model. It is further subdivided into two sub layers of protocols these are the medium access control (MAC) protocol and the logical link control (LLC) protocol. The multiple access control protocol is further be categorized into three broad categories which are the random access protocols the controlled access protocols and the channelization protocols. The main prominent random-access protocols are ALOHA, CSMA, CSMA/CD, CSMA/CA. The two approaches of Carrier Sense Multiple Access are with collision avoidance and collision detection.

Keywords

Data Link Control: It is a reliable channel for transmitting data over a dedicated link using various techniques such as framing, error control and flow control of data packets in the computer network.

Medium Access Control: It is a sublayer of the data link layer of the open system interconnections (OSI) reference model for data transmission. It is responsible for flow control and multiplexing for transmission medium. It controls the transmission of data packets via remotely shared channels.

Random Access Protocol: In this protocol, all the station has the equal priority to send the data over a channel. Depending on the channel's state (idle or busy), each station transmits the data frame. However, if more than one station sends the data over a channel, there may be a collision or data conflict.

Pure Aloha Protocol: In this protocol, when each station transmits data to a channel without checking whether the channel is idle or not, the chances of collision may occur, and the data frame can be lost. When any station transmits the data frame to a channel, the pure Aloha waits for the receiver's acknowledgment.

Slotted Aloha Protocol: In slotted Aloha, the shared channel is divided into a fixed time intervals (slots). So that, if a station wants to send a frame to a shared channel, the frame can only be sent at the beginning of the slot, and only one frame is allowed to be sent to each slot.

1-Persistent: In the 1-Persistent mode of CSMA that defines each node, it first senses the shared channel and if the channel is idle, it immediately sends the data. Else it must wait and keep track of the status of the channel to be idle and broadcast the frame unconditionally as soon as the channel is idle.

Non-Persistent: In this method, each node must sense the channel, and if the channel is inactive, it immediately sends the data. Otherwise, the station must wait for a random time (not continuously), and when the channel is found to be idle, it transmits the frames.

P-Persistent: It is the combination of 1-Persistent and Non-persistent modes. The P-Persistent mode defines that each node senses the channel, and if the channel is inactive, it sends a frame with a P

probability. If the data is not transmitted, it waits for a ($q = 1-p$ probability) random time and resumes the frame with the next time slot.

O- Persistent: In this method, the superiority of the station is defined before the transmission of the frame on the shared channel. If it is found that the channel is inactive, each station waits for its turn to retransmit the data.

Self Assessment

Fill in the blanks:

- In _____, the stations share the bandwidth of the channel in time.
a) FDMA b) CDMA c) _____ d) None of the given choices
- In _____, the chance of collision can be reduced if a station senses the medium before trying to use it.
a) _____ b) MA c) CDMA d) FDMA
- In _____ the available bandwidth is divided into frequency bands.
a) _____ b) TDMA c) CDMA d) None of the given choices.
- _____ is a multiple-access method in which the available bandwidth of a link is shared in time, frequency, or through code, between different stations.

Select the correct answer for the following questions

- Which of the following is not a Random-Access Protocol?
a) Aloha b) CDMA c) CSMA/CD d) CSMA/CA
- Which of the following is not a Controlled-Access Protocol?
a) Reservation b) Token Passing c) Pooling d) TDMA
- Which of the following is not a channelization protocol?
a) CDMA b) CSMA c) FDMA d) TDMA
- Which of the following statement is true regarding the Random-Access protocols of the Data Link Layer?
a) Here no station is superior to another station and none is assigned the control over another.
b) No station permits, or does not permit, another station to send.
c) At each instance, a station that has data to send uses a procedure defined by the protocol to make a decision on whether or not to send.
d) All of the given choices
- In Pure ALOHA, some of the frames collide because_____
a) Each frame has its individual channel and limited resources
b) Multiple channels are available for the frames
c) Multiple frames are in contention for the shared channel.
d) None of the given choices
- Which of the following is true?
a) CSMA can reduce the possibility of collision and eliminates it.
b) The possibility of collision does not exist because of propagation delay.
c) A station may sense the medium and find it idle, only because the first bit sent by another station has not yet been received.
d) None of the given choices
- Different CSMA methods that determine_____
a) What a station should do when the medium is idle?
b) What a station should do when the medium is busy?
c) Both given choices

- d) None of the given choices
12. Which of the following relates of 1-persistent method?
- After the station finds the line idle, it sends its frame immediately (with probability 1)
 - A station that has a frame to send senses the line. If the line is not idle, it waits a random amount of time and then senses the line again.
 - A station that has a frame to send senses the line. If the line is idle, it sends immediately.
 - None of the given choices
13. CSMA (Carrier Sense Multiple Access) is_____.
- a method of determining which device has access to the transmission medium at any time
 - a method access control technique for multiple-access transmission media.
 - a very common bit-oriented data link protocol issued by ISO.
 - a network access standard for connecting stations to a circuit-switched network.
14. What are the advantages of Carrier Sensing?
- Sensing a carrier can avoid simultaneous transmissions by other nodes
 - It will help in reducing retransmission of data frames due to collisions with other station data frames
 - It improves performance of the network as the individual nodes sense for a carrier present in the transmission medium before requesting access to it.
 - All the given choices
15. In Pure Aloha the vulnerable time is _____ the frame transmission time.
- The same as
 - Two times
 - Three times
 - None of the given choices

Review Questions

- How can a collision be avoided in CSMA/CD network?
- Compare and contrast CSMA/CD and token passing access methods.
- Is Slotted Aloha always better than Aloha? Explain your answer with justification.
- Explain the two functionality-oriented sublayers of the Data Link Layer.
- List the various Multiple Access Protocols and explain the various Random Access Protocols.

Answers: Self Assessment

1.	TDMA	2.	CSMA	3.	FDMA	4.	Channelization	5.	B
6.	D	7.	B	8.	D	9.	C	10.	C
11.	C	12.	A	13.	B	14.	D	15.	B

Further Readings



Andrew S. Tanenbaum, *Computer Networks*, Prentice Hall.

Behrouz A. Forouzan and Sophia Chung Fegan, *Data Communications and Networking*, McGraw-Hill Companies.

J. D. Spragins, *Telecommunications Protocols and Design*, Addison-Wesley.



<https://www.geeksforgeeks.org/multiple-access-protocols-in-computer-network>

Unit 08 : Network layer - Logical Addressing

CONTENTS

Objectives

Introduction

- 8.1 What is an IP?
- 8.2 Types of IP Addresses
- 8.3 Address Format
- 8.4 Differences between IPv4 and IPv6
- 8.5 Classful Addressing
- 8.6 What is a Mask?
- 8.7 Subnetting and Supernetting
- 8.8 Network Address Translation (NAT)
- 8.9 Address Resolution Protocol (ARP)
- 8.10 . IP Addressing

Summary

Self Assessment

Answer for self Assessment:

Further Readings

Objectives

After this lecture, you would be able to

- learn the IPv4 Datagram format and understand IPv4 addresses and classes.
- understand the need to adopt IPv6 Datagram formats.
- learn the 8-bit octet representation.
- identify the class of an IP address
- understand masks and how to use them
- learn and explore the need for IP addressing.
- understand how subnetting is used in combination with IP addressing to develop several logical addresses that exist within a single network.
- understand the concept of Public and Private IP addresses.
- learn the concept of Network Address Translation.
- understand the use of ARP and RARP

Introduction

The internet protocol is a part of the TCP IP protocol suite. The internet protocol is essential in assigning a unique IP address to a machine connected to the internet. Initially the IP addresses came in the form of IPv4 which was a 32 bit logical address which was represented as four octets of decimal numbers separated by colons these decimal numbers could take values between 0 to 255. However with extensive usage the maximum limit of the IP addresses is about to be reached. So to fulfill the need of a broader range for identifying numerous communication devices connected to the internet IPv6 was devised. It is a 128 bit hexadecimal address which can support up to 340 undecillion addresses. Further to support more number of computers in a network with limited IP addresses the concept of subnetting can overcome this limitation.

8.1 What is an IP?

An IP stands for internet protocol. An IP address is assigned to each device connected to a network. Each device uses an IP address for communication. It also behaves as an identifier as this address is used to identify the device on a network. It defines the technical format of the packets. Mainly, both the networks, i.e., IP and TCP, are combined together, so together, they are referred to as a TCP/IP. It creates a virtual connection between the source and the destination. We can also define an IP address as a numeric address assigned to each device on a network. An IP address is assigned to each device so that the device on a network can be identified uniquely. To facilitate the routing of packets, TCP/IP protocol uses a 32-bit logical address known as IPv4 (Internet Protocol version 4). An IP address consists of two parts, i.e., the first one is a network address, and the other one is a host address.

8.2 Types of IP Addresses

There are two types of IP addresses:

- a. IPv4
- b. IPv6

a. What is IPV4?

IPv4 is a version 4 of IP. It is a current version and the most commonly used IP address. It is a 32-bit address written in four numbers separated by 'dot', i.e., periods. This address is unique for each device.



Example

66.94.29.13

The above example represents the IP address in which each group of numbers separated by periods is called an Octet. Each number in an octet is in the range from 0-255. This address can produce 4,294,967,296 possible unique addresses.

In today's computer network world, computers do not understand the IP addresses in the standard numeric format as the computers understand the numbers in binary form only. The binary number can be either 1 or 0. The IPv4 consists of four sets, and these sets represent the octet. The bits in each octet represent a number. Each bit in an octet can be either 1 or 0. If the bit is 1, then the number it represents will count, and if the bit is 0, then the number it represents does not count.

Representation of 8 Bit Octet

128	64	32	16	8	4	2	1
-----	----	----	----	---	---	---	---

Figure 7.1 The above representation shows the structure of 8-bit octet.

Now, we will see how to obtain the binary representation of the above IP address, i.e., 66.94.29.13

Representation of 8 Bit Octet

Step 1: First, we find the binary number of 66.

128	64	32	16	8	4	2	1
0	1	0	0	0	0	1	0

Figure 7.2 Binary equivalent of 66

To obtain 66, we put 1 under 64 and 2 as the sum of 64 and 2 is equal to 66 ($64+2=66$), and the remaining bits will be zero, as shown above. Therefore, the binary bit version of 66 is 01000010.

Step 2: Now, we calculate the binary number of 94.

128	64	32	16	8	4	2	1
0	1	0	1	1	1	1	0

Figure 7.3 Binary equivalent of 94

To obtain 94, we put 1 under 64, 16, 8, 4, and 2 as the sum of these numbers is equal to 94, and the remaining bits will be zero. Therefore, the binary bit version of 94 is 01011110.

Step 3: The next number is 29.

128	64	32	16	8	4	2	1
0	0	0	1	1	1	0	0

Figure 7.4 Binary equivalent of 29

To obtain 29, we put 1 under 16, 8, 4, and 1 as the sum of these numbers is equal to 29, and the remaining bits will be zero. Therefore, the binary bit version of 29 is 00011101.

Step 4: The last number is 13.

128	64	32	16	8	4	2	1
0	0	0	0	1	1	0	1

Figure 7.5 Binary equivalent of 13

To obtain 13, we put 1 under 8, 4, and 1 as the sum of these numbers is equal to 13, and the remaining bits will be zero. Therefore, the binary bit version of 13 is 00001101.

Drawback of IPv4

Currently, the population of the world is 7.6 billion. Every user is having more than one device connected with the internet, and private companies also rely on the internet. As we know that IPv4 produces four billion addresses, which are not enough for each device connected to the internet on a planet? Although the various techniques were invented, such as variable-length mask, network address translation, port address translation, classes, inter-domain translation, to conserve the bandwidth of IP address and slow down the depletion of an IP address. In these techniques, public IP is converted into a private IP due to which the user having public IP can also use the internet. But still, this was not so efficient, so it gave rise to the development of the next generation of IP addresses, i.e., IPv6.

b. What is IPv6?

IPv4 produces 4 billion addresses, and the developers think that these addresses are enough, but they were wrong. IPv6 is the next generation of IP addresses. The main difference between IPv4 and IPv6 is the address size of IP addresses. The IPv4 is a 32-bit address, whereas IPv6 is a 128-bit hexadecimal address. IPv6 provides a large address space, and it contains a simple header as compared to IPv4.

It provides transition strategies that convert IPv4 into IPv6, and these strategies are as follows:

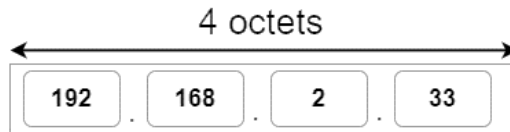
- **Dual stacking:** It allows us to have both the versions, i.e., IPv4 and IPv6, on the same device.
- **Tunneling:** In this approach, all the users have IPv6 communicates with an IPv4 network to reach IPv6.
- **Network Address Translation:** The translation allows the communication between the hosts having a different version of IP.

This hexadecimal address contains both numbers and alphabets. Due to the usage of both the numbers and alphabets, IPv6 is capable of producing over 340 undecillion (3.4×10^{38}) addresses.

IPv6 is a 128-bit hexadecimal address made up of 8 sets of 16 bits each, and these 8 sets are separated by a colon. In IPv6, each hexadecimal character represents 4 bits. So, we need to convert 4 bits to a hexadecimal number at a time

8.3 Address Format

The address format of IPv4:



The address format of IPv6:

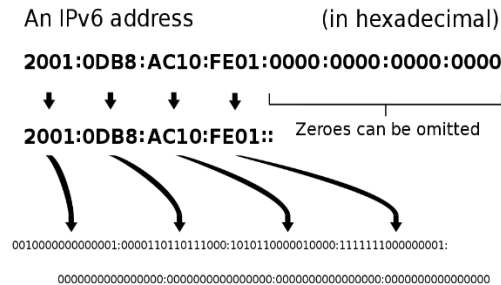


The above diagram shows the address format of IPv4 and IPv6. An IPv4 is a 32-bit decimal address. It contains 4 octets or fields separated by 'dot', and each field is 8-bit in size. The number that each field contains should be in the range of 0-255. Whereas an IPv6 is a 128-bit hexadecimal address. It contains 8 fields separated by a colon, and each field is 16-bit in size.

8.4 Differences between IPv4 and IPv6

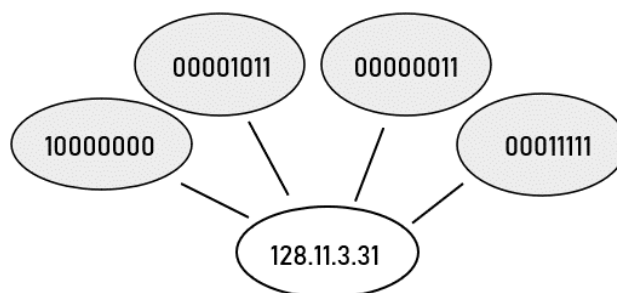
	IPv4	IPv6
Address Length	IPv4 is a 32-bit address.	IPv6 is a 128-bit address.
Fields	IPv4 is a numeric address that consists of 4 fields which are separated by dot (.).	IPv6 is an alphanumeric address that consists of 8 fields which are separated by colon.
Classes	IPv4 has 5 different classes of IP address that includes Class A, Class B, Class C, Class D and Class E	IPv6 does not contain classes of IP addresses.
Number of IP address	IPv4 has a limited number of IP addresses.	IPv6 has a large number of IP addresses
VLSM	It supports VLSM (Virtual Length Subnet Mask). Here VLSM means that IPv4 converts IP addresses into a subnet of different sizes.	It does not support VLSM.
Address Configuration	It supports manual and DHCP configuration.	It supports manual, DHCP, auto-configuration, and renumbering.
Address Space	It generates 4 billion unique addresses	It generates 340 undecillion unique addresses.
End-to-end connection integrity	In IPv4, end-to-end connection integrity is unachievable	In case of IPv6, end-to-end connection integrity is achievable
Security features	In IPv4, security depends on the application. This IP address is not developed in keeping the security features in mind.	In IPv6, IPSEC is developed for security purposes.
Address representation	In IPv4, the IP address is represented in decimal.	In IPv6 representation of the IP address in hexadecimal.
Fragmentation	Fragmentation is done by the sender and the forwarding routers.	Fragmentation is done by the senders only.
Packet flow identification	It does not provide any mechanism for packet flow identification.	It uses flow label field in the header for the packet flow identification.
Checksum field	The checksum field is available in IPv4.	The checksum field is not available in IPv6.

Transmission Scheme	IPv4 is broadcasting.	On the other hand, IPv6 is multicasting, which provides efficient network operations.
Encryption and Authentication	It does not provide encryption and authentication.	It provides encryption and authentication.
Number of octets	It consists of 4 octets.	It consists of 8 fields and each field contains 2 octets. Therefore, the total number of octets in IPv6 is 16.



1. About IPV6

The identifier used in the IP layer of the TCP/IP protocol suite to identify each device connected to the Internet is called the Internet address or IP address. An IP address is a 32-bit address that uniquely and universally defines the connection of a host or a router to the Internet. IP addresses are unique. They are unique in the sense that each address defines one, and only one, connection to the Internet. Two devices on the Internet can never have the same address. An IP address is a 32-bit address. The IP addresses are unique. The address space of IPv4 is 2³² or 4,294,967,296.



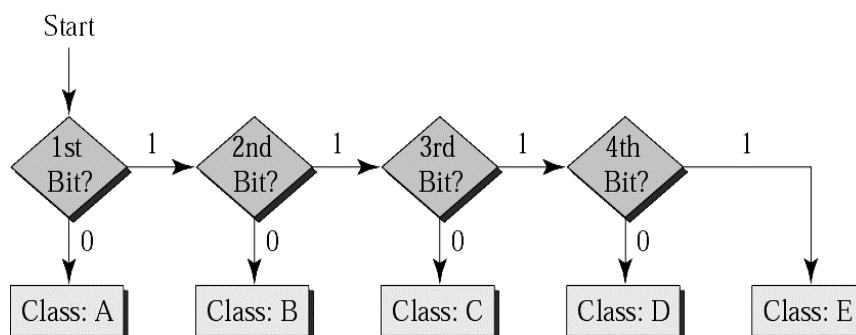
8.5 Classful Addressing

IP addresses, when started a few decades ago, used the concept of classes. This architecture is called classful addressing. In the mid-1990s, a new architecture, called classless addressing, was introduced and will eventually supersede the original architecture. However, part of the Internet is still using classful addressing, but the migration is very fast.

Finding the class in binary notation

	First byte	Second byte	Third byte	Fourth byte
Class A	0			
Class B	10			
Class C	110			
Class D	1110			
Class E	1111			

Figure:

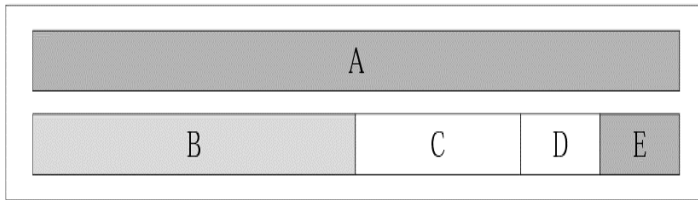


	First byte	Second byte	Third byte	Fourth byte
Class A	0 to 127			
Class B	128 to 191			
Class C	192 to 223			
Class D	224 to 239			
Class E	240 to 255			

1. Occupation of the address space

Class	Number of Addresses	Percentage
A	$2^{31} = 2,147,483,648$	50%
B	$2^{30} = 1,073,741,824$	25%
C	$2^{29} = 536,870,912$	12.5%
D	$2^{28} = 268,435,456$	6.25%
E	$2^{28} = 268,435,456$	6.25%

Address space



2. Class Wise IP Address Ranges

CLASS	LEADING BITS	NET ID BITS	HOST ID BITS	NO. OF NETWORKS	ADDRESSES PER NETWORK	START ADDRESS	END ADDRESS
CLASS A	0	8	24	2^7 (128)	2^{24} (16,777,216)	0.0.0.0	127.255.255.255
CLASS B	10	16	16	2^{14} (16,384)	2^{16} (65,536)	128.0.0.0	191.255.255.255
CLASS C	110	24	8	2^{21} (2,097,152)	2^8 (256)	192.0.0.0	223.255.255.255
CLASS D	1110	NOT DEFINED	NOT DEFINED	NOT DEFINED	NOT DEFINED	224.0.0.0	239.255.255.255
CLASS E	1111	NOT DEFINED	NOT DEFINED	NOT DEFINED	NOT DEFINED	240.0.0.0	255.255.255.255

3. Class Wise IP Address Ranges

Class A addresses were designed for large organizations with a large number of attached hosts or routers.

Class B addresses were designed for midsize organizations with tens of thousands of attached hosts or routers.

Class C addresses were designed for small organizations with a small number of attached hosts or routers

All the network bits set as 1

and all the host bits set as 0

Subnet Mask

11111111.11111111.00000000.00000000

255 . 255 . 0 . 0



How can we prove that we have 2,147,483,648 addresses in class A?

Example:

Solution

In class A, only 1 bit defines the class. The remaining 31 bits are available for the address. With 31 bits, we can have 2^{31} or 2,147,483,648 addresses.



Find the class of the address 00000001 00001011 00001011 11101111

Example:

Solution

The first bit is 0. This is a class A address.



Example:

Find the class of the IP address 227.12.14.87

Solution

The first byte is 227 (between 224 and 239); the class is D.

4. Netid and Hostid

In classful addressing, an IP address in class A, B, or C is divided into netid and hostid. These parts are of varying lengths, depending on the class of the address. Note that the concept does not apply to classes D and E. In class A, one byte defines the netid and three bytes define the hostid. In class B, two bytes define the netid and two bytes define the hostid. In class C, three bytes define the netid and one byte defines the hostid.

8.6 What is a Mask?

Although the length of the netid and hostid (in bits) is predetermined in classful addressing, we can also use a mask (also called the default mask), a 32-bit number made of contiguous 1s. The mask can help us to find the netid and the hostid.



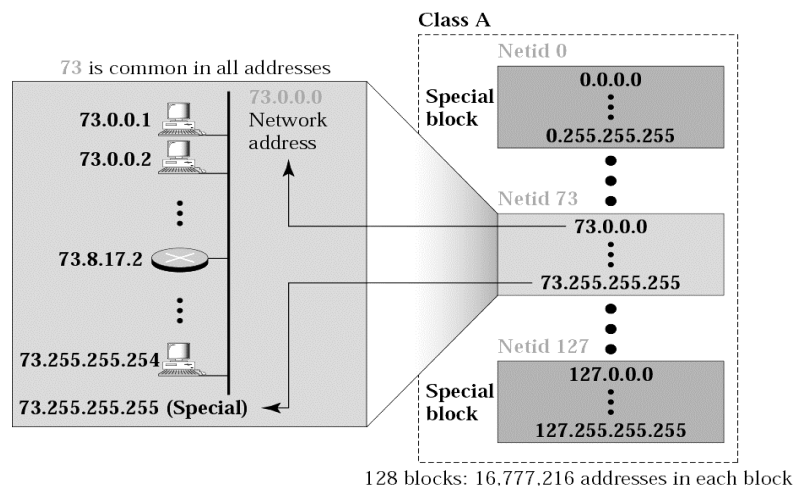
Example:

The mask for a class A address has eight 1s, which means the first 8 bits of any address in class A define the netid; the next 24 bits define the hostid.

Default masks

Class	Mask in Binary	Mask in dotted-decimal
A	11111111 00000000 00000000 00000000	255.0.0.0
B	11111111 11111111 00000000 00000000	255.255.0.0
C	11111111 11111111 11111111 00000000	255.255.255.0

Blocks in class A

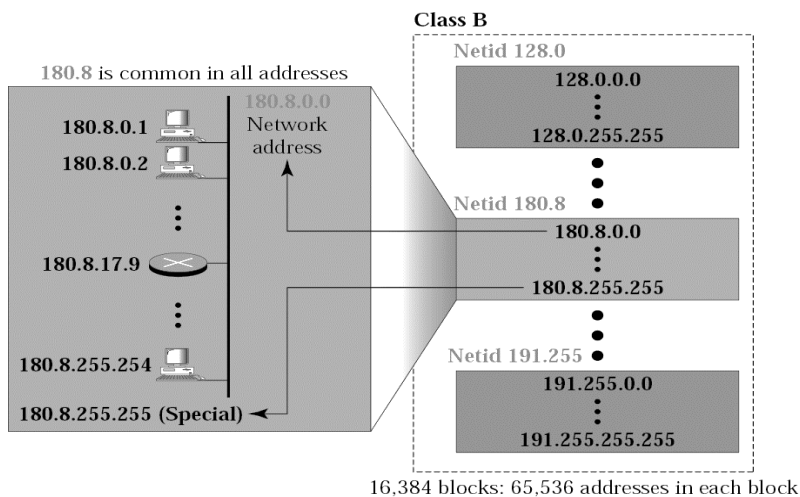


128 blocks: 16,777,216 addresses in each block



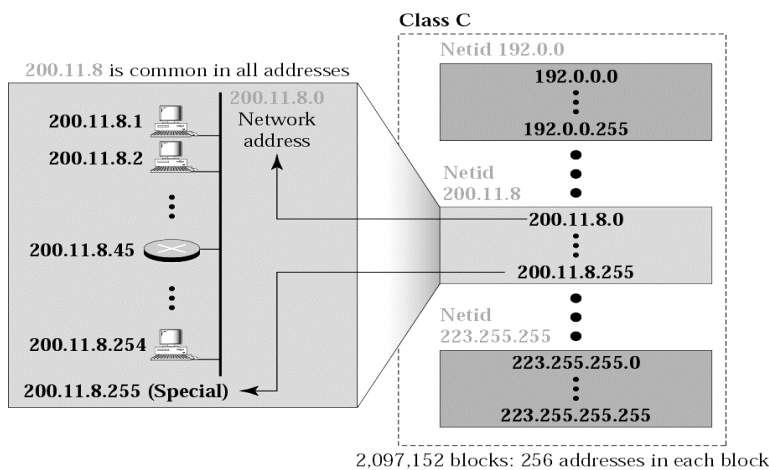
Million of class A addresses are wasted

Blocks in Class B



Million of class A addresses are wasted

Blocks in Class C



The number of addresses in class C is smaller than the needs of most organizations. Class D addresses are used for multicasting; there is only one block in this class. In classful addressing, the network address (the first address in the block) is the one that is assigned to the organization. The range of addresses can automatically be inferred from the network address. The network address is the beginning address of each block. It can be found by applying the default mask to any of the addresses in the block (including itself). It retains the netted of the block and sets the hosted to zero



Example:

Given the address 132.6.17.85, find the beginning address (network address).

Solution

The default mask is 255.255.0.0, which means that the first 2 bytes are preserved, and the other 2 bytes are set to 0s. The network address is 132.6.0.0.



Example:

Given the address 201.180.56.5, find the beginning address (network address).

Solution

The default mask is 255.255.255.0, which means that the first 3 bytes are preserved, and the last byte is set to 0. The network address is 201.180.56.0.



We must not apply the default mask of one class to an address belonging to another class

8.7 Subnetting and Supernetting

The main problem with classful addressing is that the network addresses available for assignment to organizations are close to depletion. Also there is an ever-increasing demand for addresses from organizations that want connection to the Internet. So there are two solutions: sub netting and super netting.



IP addresses are designed with two levels of hierarchy

Need for IP Addressing and Subnetting in Computer Networks. IP addressing is used to recognize the host of a network and uniquely identify a particular device of the Network. Whereas subnetting is used in combination with IP addressing to develop several logical addressing that exists within a single network. Need for IP Addressing and Subnetting in Computer Networks

In our daily life, we human beings identify each other with our names, similarly, the routers and switches recognize their neighboring device and network with an IP address and a subnet mask.

Subnetting

Subnetting allows us to create various sub-networks or logical networks within one network of a particular class of the network. Without subnetting, it is almost unrealistic to create big networks. For constructing a big networking system, every link must have a unique IP address with every device on that linked network which is being the participant of that network. With the help of a subnetting technique, we can split the large networks of a particular class (A, B or C) into smaller subnet works for inter-connection between each node which are situated at different locations. Each node on the network would have a distinctive IP and subnet mask IP. Any switch, router or gateway that connects 'n' networks has 'n' unique Network ID's and one Subnet Mask for each of the network it is interconnecting with.

The formulae of subnetting is $2^n \geq \text{requirement}$, where n is the number of networks.

The formulae of a number of hosts per subnet is: $2^n - 2$



Example: Now let's understand the overall process with the help of an example. We have taken an example of Class C network ID with a default subnet mask. Suppose Network ID/IP address is: 192.168.1.0. Default Subnet mask: 255.255.255.0 (in decimal)

Default Subnet mask: 11111111.11111111.11111111.00000000 (in binary). Thus, the numbers of bits are $8+8+8+0 = 24$ bits. As mentioned earlier, for subnetting in class C network, we will borrow bits from the host portion of the subnet mask. Therefore to customize the subnet as per requirement. We take a subnet mask of 255.255.255.248 (in decimal) which is equivalent to 11111111.11111111.11111111.11110000 (in binary).

From the above binary notation, we can see that the last 3 bits of the last octet can be used for host ID addressing purpose.

Thus the number of subnets = $2^n = 2^3 = 8$ subnets ($n=3$).

Number of hosts per subnet = $2^n - 2 = 2^3 - 2 = 8 - 2 = 6$ Subnets i.e. usable Host IP.

Now the IP addressing scheme is as follows:

Network IP	First Usable IP	Last Usable IP	Broadcast IP
192.168.1.0	192.168.1.1	192.168.1.6	192.168.1.7
192.168.1.8	192.168.1.9	192.168.1.14	192.168.1.15
192.168.1.16	192.168.1.17	192.168.1.22	192.168.1.23
192.168.1.24	192.168.1.25	192.168.1.30	192.168.1.31
192.168.1.32	192.168.1.33	192.168.1.38	192.168.1.39
192.168.1.40	192.168.1.41	192.168.1.46	192.168.1.47
192.168.1.48	192.168.1.49	192.168.1.54	192.168.1.55
192.168.1.56	192.168.1.57	192.168.1.62	192.168.1.63

The subnet mask for all the above IP's in the table is common i.e. 255.255.255.248.

With the help of this example, we can clearly see, how subnetting helps us to construct inter-networking between various links and nodes of the same sub network. All these above IP's can be used for inter-networking the devices within the overall network.



that the subnet mask is most widely used everywhere in a computer networking system. Hence, there is one more method to represent the subnet mask of a particular network which is chosen and standardized as it is easy to denote and memorize.

Different Subnetting Schemes in Binary and Decimal

Subnet mask- 255.255.255.248 (binary) which is equivalent to 11111111.11111111.11111111.11110000 (decimal notation). From the decimal notation we can calculate the number of bits having 1 in each octet: $8+8+8+5=29$

Thus the Subnet mask can be denoted as /29.

With Network ID it can be denoted as 192.168.1.9/29.

From the above notation, anyone who knows the standard notation and formulae of subnetting can understand that the IP is using a subnet mask of 255.255.255.248 or /29.

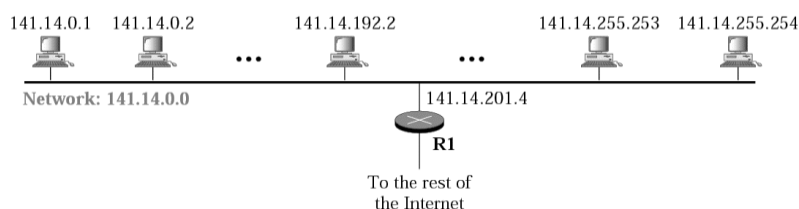


Example: The different subnetting scheme in binary and decimal notation is shown below:

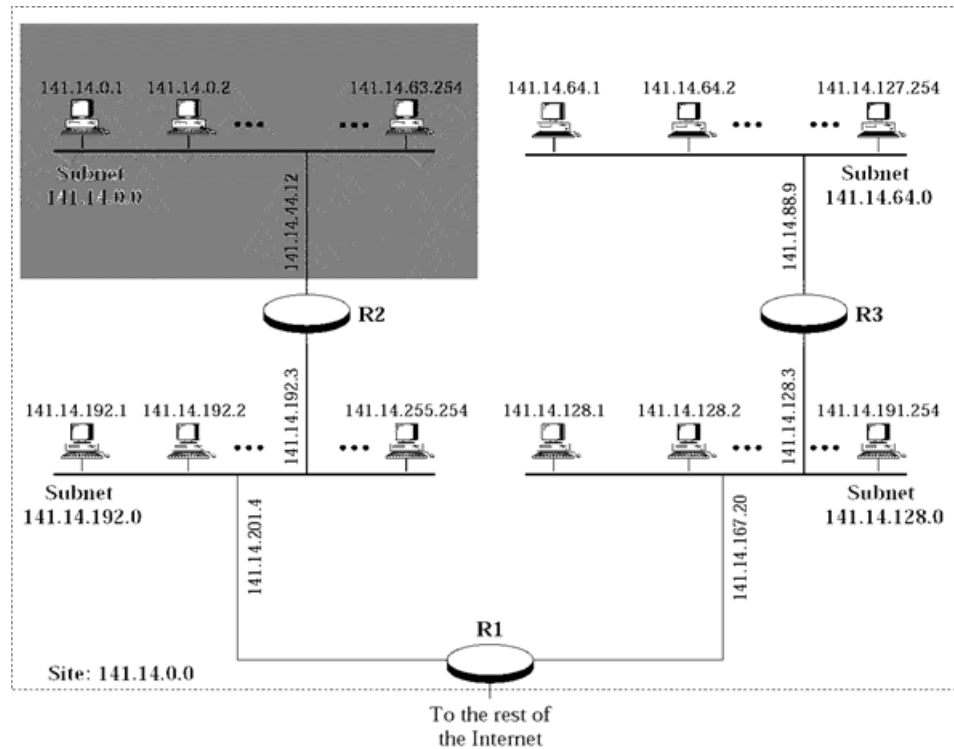
Subnet Mask	Notation in decimal	Notation in Binary	Number of Usable IP
/24	255.255.255.0	11111111.11111111.11111111.00000000	254
/25	255.255.255.128	11111111.11111111.11111111.10000000	126
/26	255.255.255.192	11111111.11111111.11111111.11000000	62
/27	255.255.255.224	11111111.11111111.11111111.11100000	30
/28	255.255.255.240	11111111.11111111.11111111.11110000	14
/29	255.255.255.248	11111111.11111111.11111111.11111000	6
/30	255.255.255.252	11111111.11111111.11111111.11111100	2

The '/' notation method of the subnet mask is most widely used as it is easy to memorize and the binary notation and decimal are very lengthy in size. As we are denoting the mask scheme while interconnecting the network components through the figure, if we use the decimal and binary method then the overall diagram will become very complex and difficult to understand. There are so many IP's on the platform to be shown and it becomes difficult to memorize as well. Thus generally, people who are familiar with routing and IP addressing scheme use short notation methods in figures and diagrams.

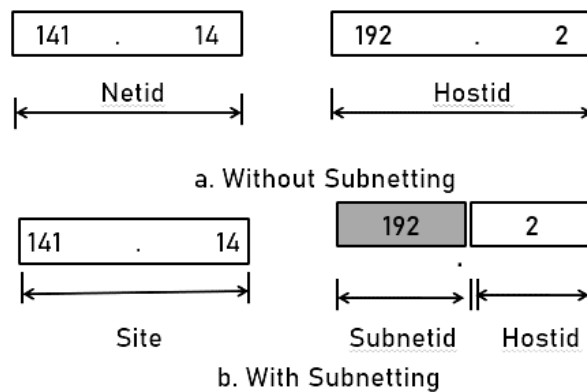
A Network with Two Levels of Hierarchy (not Sub netted)



A network with three levels of hierarchy (sub netted)



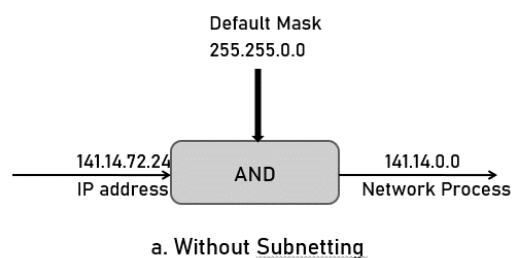
Addresses in a Network with and without Subnetting

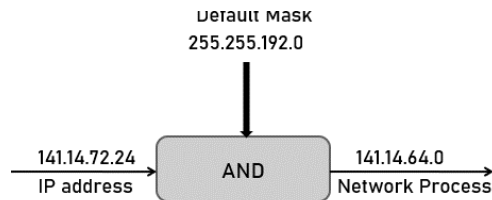


Hierarchy concept in a telephone number



Default mask and subnet mask





b. With Subnetting



Example: What is the sub network address if the destination address is 200.45.34.56 and the subnet mask is 255.255.240.0?

Solution

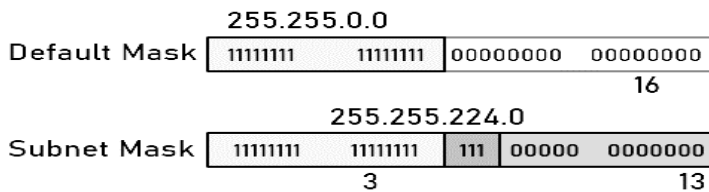
We apply the AND operation on the address and the subnet mask.

Address 11001000 00101101 00100010 00111000

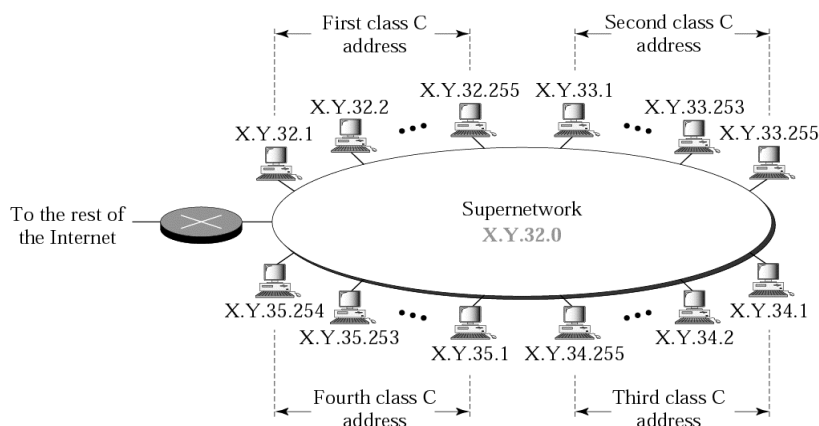
Subnet Mask 11111111 11111111 11110000 00000000

Subnetwork Address 11001000 00101101 00100000 00000000.

1. Comparison of a Default Mask and a Subnet mask

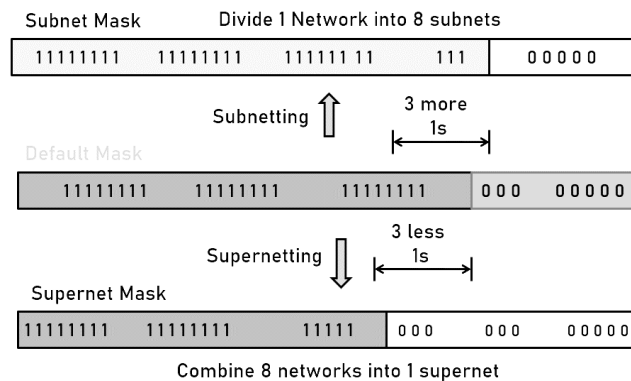


Example of Super network



In subnetting, we need the first address of the subnet and the subnet mask to define the range of addresses. In super netting, we need the first address of the super net and the super net mask to define the range of addresses.

2. Comparison of Subnet, Default, and Supernet masks



The idea of subnetting and super netting of classful addresses is almost obsolete.

8.8 Network Address Translation (NAT)

To access Internet, one public IP address is needed, but we can use a private IP address in our private network. NAT allows multiple devices to access the Internet through a single public address. To achieve this, the translation of private IP address to a public IP address is required. One or more local IP address is translated into one or more Global IP address and vice versa in order to provide Internet access to the local hosts. Also, it does the translation of port numbers i.e., masks the port number of the host with another port number, in the packet that will be routed to the destination. It then makes the corresponding entries of IP address and port number in the NAT table. NAT generally operates on router or firewall.

1. Network Address Translation -Working

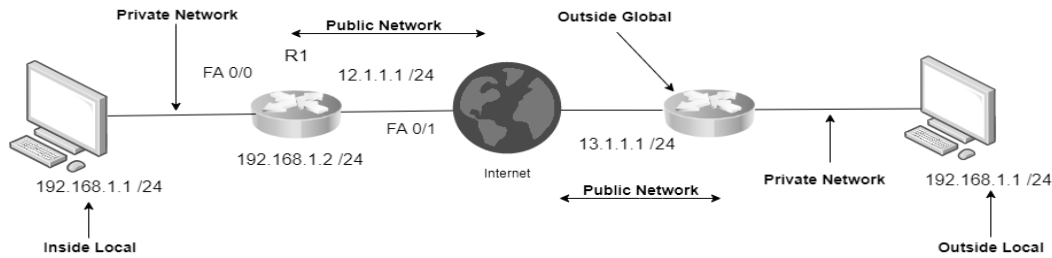
The border router is configured for NAT i.e the router which has one interface in local (inside) network and one interface in the global (outside) network. When a packet traverse outside the local network, then NAT converts that local IP address to a global IP address. When a packet enters the local network, the global IP address is converted to a local IP address. If NAT run out of addresses, then the packets will be dropped and an ICMP host unreachable packet to the destination is sent.

2. Why Mask Port Numbers?

Suppose, in a network, two hosts A and B are connected. Now, both simultaneously request for the same destination, on the same port number, say 1000, on the host side. If NAT does an only translation of IP addresses, then when their packets will arrive at the NAT, both of their IP addresses would be masked by the public IP address of the network and sent to the destination. Destination will send replies on the public IP address of the router. Thus, on receiving a reply, it will be unclear to NAT as to which reply belongs to which host (because source port numbers for both A and B are same). Destination will send replies on the public IP address of the router. Thus, on receiving a reply, it will be unclear to NAT as to which reply belongs to which host (because source port numbers for both A and B are same). Hence, to avoid such a problem, NAT masks the source port number as well and makes an entry in the NAT table.

3. NAT Inside and Outside Addresses

Inside refers to the addresses which must be translated. Inside refers to the addresses which must be translated. Outside refers to the addresses which are not in control of an organization. Inside refers to the addresses which must be translated. Outside refers to the addresses which are not in control of an organization. These are the network Addresses in which the translation of the addresses will be done.



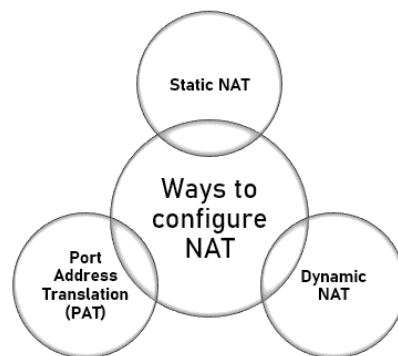
4. Inside Local Address

An IP address that is assigned to a host on the Inside (local) network. The address is probably not a IP address assigned by the service provider i.e., these are private IP address. This is the inside host seen from the inside network. IP address that represents one or more inside local IP addresses to the outside world. This is the inside host as seen from the outside network.

5. Outside Local Address

This is the actual IP address of the destination host in the local network after translation. This is the outside host as seen from the outside network. It is the IP address of the outside destination host before translation.

Network Address Translation Types



6. Static NAT

In this, a single unregistered (Private) IP address is mapped with a legally registered (Public) IP address i.e one-to-one mapping between local and global address. This is generally used for Web hosting. These are not used in organizations as there are many devices who will need Internet access and to provide Internet access, the public IP address is needed. Suppose, if there are 3000 devices who need access to the Internet, the organization must buy 3000 public addresses that will be very costly.

7. Dynamic NAT

Here, an unregistered IP address is translated into a registered (Public) IP address from a pool of public IP address. If the IP address of pool is not free, then the packet will be dropped as only a fixed number of private IP address can be translated to public addresses. Suppose, if there is a pool of 2 public IP addresses then only 2 private IP addresses can be translated at a given time. If 3rd private IP address wants to access Internet then the packet will be dropped therefore many private IP addresses are mapped to a pool of public IP addresses. NAT is used when the number of users who wants to access the Internet is fixed. This is also very costly as the organization must buy many global IP addresses to make a pool.

8. Port Address Translation (PAT)

This is also known as NAT overload. In this, many local (private) IP addresses can be translated to a single registered IP address. Port numbers are used to distinguish the traffic i.e., which traffic belongs to which IP address. NAT conserves legally registered IP addresses.

9. Network Address Translation- Advantages

Being cost-effective, it is most frequently used as thousands of users can be connected to the Internet by using only one real global (public) IP address. It provides privacy as the device IP address, sending and receiving the traffic, will be hidden. Eliminates address renumbering when a network evolves. Translation results in switching path delays. Certain applications will not function while NAT is enabled. Complicates tunneling protocols such as IPsec.

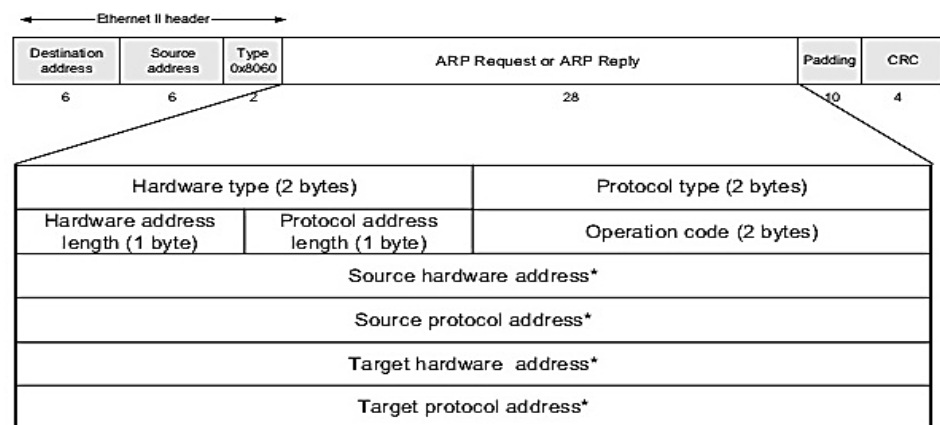
10. Network Address Translation - Disadvantages

Also, router being a network layer device should not tamper with port numbers (transport layer) but it has to do so because of NAT.

8.9 Address Resolution Protocol (ARP)

Address Resolution Protocol (ARP) is a communication protocol used to find the MAC (Media Access Control) address of a device from its IP address. This protocol is used when a device wants to communicate with another device on a Local Area Network or Ethernet. At Network layer we are locating IP's of Host and sending machine, but at DLL we need to know the MAC address of the machine. By sending an ARP request packet. The ARP request will have the following structure:

The structure of a typical ARP request packet is:



* Note: The length of the address fields is determined by the corresponding address length fields

It will be broadcasted to all but will be replied only by the correct IP as it the Destination IP Address mentioned in the request

Its just like a teacher asking a question in the class, then only the student whose name is broadcasted will reply in unicast form to the question. Reply to this is

Destination		Source	
Destination IP Address	Destination MAC Address ?	Source MAC Address	Since <i>destination MAC address is not known</i> so we will put Broadcast Address

It will be broadcasted to all but will be replied only by the correct IP as it the Destination IP Address mentioned in the request. Its just like a teacher asking a question in the class, then only the student whose name is broadcasted will reply in unicast form to the question. Reply to this is

IP address	MAC Address	MAC of Destination	MAC of Host
------------	-------------	--------------------	-------------

1. Types of ARP

There are four types of Address Resolution Protocol, which is given below:

- Proxy ARP
- Gratuitous ARP
- Reverse ARP (RARP)

- Inverse ARP



2. Proxy ARP

To summarize we can say: Proxy ARP is a method through which a Layer 3 devices may respond to ARP requests for a target that is in a different network from the sender. The Proxy ARP configured router responds to the ARP and map the MAC address of the router with the target IP address and fool the sender that it is reached at its destination.

At the backend, the proxy router sends its packets to the appropriate destination because the packets contain the necessary information.



Example: If Host A wants to transmit data to Host B, which is on the different network, then Host A sends an ARP request message to receive a MAC address for Host B. The router responds to Host A with its own MAC address pretend itself as a destination. When the data is transmitted to the destination by Host A, it will send to the gateway so that it sends to Host B. This is known as proxy ARP.

3. Gratuitous ARP

It is an ARP request of the host that helps to identify the duplicate IP address. It is a broadcast request for the IP address of the router. If an ARP request is sent by a switch or router to get its IP address and no ARP responses are received, so all other nodes cannot use the IP address allocated to that switch or router. Yet if a router or switch sends an ARP request for its IP address and receives an ARP response, another node uses the IP address allocated to the switch or router.



There are some primary use cases of gratuitous ARP that are given below:

The gratuitous ARP is used to update the ARP table of other devices. It also checks whether the host is using the original IP address or a duplicate one.

4. Reverse ARP (RARP)

It is a networking protocol used by the client system in a local area network (LAN) to request its IPv4 address from the ARP gateway router table. A table is created by the network administrator in the gateway-router that is used to find out the MAC address to the corresponding IP address. When a new system is set up or any machine that has no memory to store the IP address, then the user has to find the IP address of the device. The device sends a RARP broadcast packet, including its own MAC address in the address field of both the sender and the receiver hardware. A host installed inside of the local network called the RARP-server is prepared to respond to such type of broadcast packet. The RARP server is then trying to locate a mapping table entry in the IP to MAC address. If any entry matches the item in the table, then the RARP server sends the response packet along with the IP address to the requesting computer.

5. Inverse ARP (in ARP)

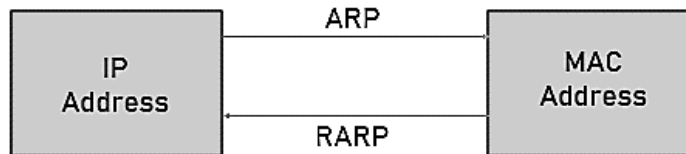
Inverse ARP is Inverse of the ARP, and it is used to find the IP addresses of the nodes from the data link layer addresses. These are mainly used for the frame relays, and ATM networks, where Layer 2 virtual circuit addressing are often acquired from Layer 2 signaling. When using these virtual circuits, the relevant Layer 3 addresses are available. ARP conversions Layer 3 addresses to Layer 2

addresses. However, its opposite address can be defined by in ARP. The in ARP has a similar packet format as ARP, but operational codes are different.

6. Difference between ARP and RARP

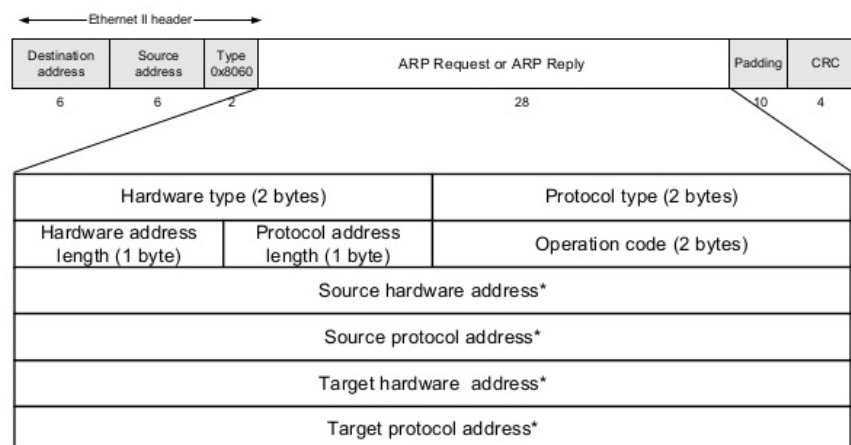
In Address Resolution Protocol (ARP), Receiver's MAC address is fetched. Through ARP, (32-bit) IP address mapped into (48-bit) MAC address. Whereas, in Reverse Address Resolution Protocol (RARP), IP address is fetched through server. Through RARP, (48-bit) MAC address of 48 bits mapped into (32-bit) IP address.

Difference between ARP and RARP



Sr. No.	ARP	RARP
1.	ARP stands for Address Resolution Protocol.	Whereas RARP stands for Reverse Address Resolution Protocol.
2.	Through ARP, (32-bit) IP address mapped into (48-bit) MAC address.	Whereas through RARP, (48-bit) MAC address of 48 bits mapped into (32-bit) IP address.
3.	In ARP, broadcast MAC address is used.	While in RARP, broadcast IP address is used.
4.	In ARP, ARP table is managed or maintained by local host.	While in RARP, RARP table is managed or maintained by RARP server.
5.	In Address Resolution Protocol, Receiver's MAC address is fetched.	While in RARP, IP address is fetched.
6.	In ARP, ARP table uses ARP reply for its updation.	While in RARP, RARP table uses RARP reply for configuration of IP addresses .
7.	Hosts and routers uses ARP for knowing the MAC address of other hosts and routers in the networks.	While RARP is used by small users having less facilities.

7. ARP Packet Format



* Note: The length of the address fields is determined by the corresponding address length fields

8.10. IP Addressing

An IP address is a unique address that identifies a device on the internet or a local network. IP stands for "Internet Protocol," which is the set of rules governing the format of data sent via the

internet or local network. They contain location information and make devices accessible for communication. In essence, IP addresses are the identifier that allows information to be sent between devices on a network. The internet needs a way to differentiate between different computers, routers, and websites. IP addresses provide a way of doing so and form an essential part of how the internet works.

1. What is an IP?

An IP address is a string of numbers separated by periods. IP addresses are expressed as a set of four numbers.



Example: An example address might be 192.158.1.38. Each number in the set can range from 0 to 255. So, the full IP addressing range goes from 0.0.0.0 to 255.255.255.255. IP addresses are not random. They are mathematically produced and allocated by the Internet Assigned Numbers Authority (IANA), a division of the Internet Corporation for Assigned Names and Numbers (ICANN).

2. How do IP Addresses Work?

Internet Protocol works the same way as any other language, by communicating using set guidelines to pass information. All devices find, send, and exchange information with other connected devices using this protocol. By speaking the same language, any computer in any location can talk to one another. The use of IP addresses typically happens behind the scenes.

The process works like this:

1. Your device indirectly connects to the internet by connecting at first to a network connected to the internet, which then grants your device access to the internet.
2. When you are at home, that network will probably be your Internet Service Provider (ISP). At work, it will be your company network.
3. Your IP address is assigned to your device by your ISP.
4. Your internet activity goes through the ISP, and they route it back to you, using your IP address. Since they are giving you access to the internet, it is their role to assign an IP address to your device.
5. However, your IP address can change. For example, turning your modem or router on or off can change it. Or you can contact your ISP, and they can change it for you.
6. When you are out and about – for example, traveling – and you take your device with you, your home IP address does not come with you. This is because you will be using another network (Wi-Fi at a hotel, airport, or coffee shop, etc.) to access the internet and will be using a different (and temporary) IP address, assigned to you by the ISP of the hotel, airport or coffee shop.

3. Types of IP Addresses

There are different categories of IP addresses, and within each category, different types.

- 1) Consumer IP addresses
 - a) Private IP addresses
 - b) Public IP addresses
- 2) Dynamic IP addresses
- 3) Static IP addresses

1. Consumer IP addresses

Every individual or business with an internet service plan will have two types of IP addresses:

Their Private IP addresses and

Their Public IP address.

The terms public and private relate to the network location. That is, a private IP address is used inside a network, while a public one is used outside a network.

a) Private IP addresses

Every device that connects to your internet network has a private IP address. This includes computers, smartphones, and tablets but also any Bluetooth-enabled devices like speakers, printers, or smart TVs. With the growing internet of things, the number of private IP addresses you have at home is probably growing. Your router needs a way to identify these items separately, and many items need a way to recognize each other. Therefore, your router generates private IP addresses that are unique identifiers for each device that differentiate them on the network.

b) Public IP addresses

A public IP address is the primary address associated with your whole network. While each connected device has its own IP address, they are also included within the main IP address for your network. As described above, your public IP address is provided to your router by your ISP. Typically, ISPs have a large pool of IP addresses that they distribute to their customers. Your public IP address is the address that all the devices outside your internet network will use to recognize your network. Public IP addresses come in two forms –

i. Dynamic and

ii. Static.

i) Dynamic IP addresses

Dynamic IP addresses change automatically and regularly. ISPs buy a large pool of IP addresses and assign them automatically to their customers. Periodically, they re-assign them and put the older IP addresses back into the pool to be used for other customers. The rationale for this approach is to generate cost savings for the ISP. Automating the regular movement of IP addresses means they don't have to carry out specific actions to re-establish a customer's IP address if they move home. For example, there are security benefits, too, because a changing IP address makes it harder for criminals to hack into your network interface.

ii) Static IP addresses

In contrast to dynamic IP addresses, static addresses remain consistent. Once the network assigns an IP address, it remains the same. Most individuals and businesses do not need a static IP address, but for businesses that plan to host their own server, it is crucial to have one. This is because a static IP address ensures that websites and email addresses tied to it will have a consistent IP address – vital if you want other devices to be able to find them consistently on the web.

For website owners who don't host their own server, and instead rely on a web hosting package – which is the case for most websites – there are two types of website IP addresses.

4. Types of IP website Addresses

i. Shared IP addresses

ii. Dedicated IP addresses

i. Shared IP addresses

Websites that rely on shared hosting plans from web hosting providers will typically be one of many websites hosted on the same server. This tends to be the case for individual websites or SME websites, where traffic volumes are manageable, and the sites themselves are limited in terms of the number of pages, etc. Websites hosted in this way will have shared IP addresses.

ii. Dedicated IP addresses

Some web hosting plans have the option to purchase a dedicated IP address (or addresses). This can make obtaining an SSL certificate easier and allows you to run your own File Transfer Protocol (FTP) server. This makes it easier to share and transfer files with multiple people within an organization and allow anonymous FTP sharing options. A dedicated IP address also allows you to access your website using the IP address alone rather than the domain name – useful if you want to build and test it before registering your domain.

5. How to Find Your Private IP Addresses?

Finding your private IP address varies by platform:

In Windows:

- Use the command prompt.
- Search for "cmd" (without the quotes) using Windows search
- In the resulting pop-up box, type "ipconfig" (no quote marks) to find the information.

On a Mac:

- Go to System Preferences
- Select network – and the information should be visible.

On an iPhone:

- Go to Settings
- Select Wi-Fi and click the "i" in a circle () next to the network you are on – the IP address should be visible under the DHCP tab.
- If you need to check the IP addresses of other devices on your network, go into the router.
- How you access the router depends on the brand and the software it uses.
- Generally, you should be able to type the router's gateway IP address into a web browser on the same network to access it.
- From there, you will need to navigate to something like "attached devices," which should display a list of all the devices currently or recently attached to the network – including their IP addresses.

6. IP Addressing

Class of IP Address	Total Number of IP Addresses	First Octet Decimal Range	Number of Networks Available	Hosts per Network	Default Subnet Mask
Class A	2^{31}	1-126	2^7-2	$2^{24}-2$	255.0.0.0
Class B	2^{30}	128-191	2^{14}	$2^{16}-2$	255.255.0.0
Class C	2^{29}	192-223	2^{21}	2^8-2	255.255.255.0
Class D	2^{28}	224-239	Not Defined	Not Defined	Not Defined
Class E	2^{28}	240-254	Not Defined	Not Defined	Not Defined

7. Class A

Starts with "0"

Total Number of IP Addresses available in Class A = Numbers possible due to remaining available 31 bits $=2^{31}$		Total number of Hosts that can be configured in Class A = Numbers possible due to available 24 bits in the Host Id -2 $=2^{24}-2$	
Net ID(8) Host ID(24) 32 bits		Range of 1 st Octet Minimum value of 1 st Octet = 00000000 = 0 Range=[0,127] But because 2 networks are reserved and	

	unused, Range of 1 st octet=[1,126]
Total number of networks available in Class A = numbers possible due to remaining available 7 bits in the Net ID -2 = 2^7-2	Class A is used by organizations requiring very large size networks like NASA, Pentagon etc.

In Class A, total number of IP Addresses available for networks are 2 less.

This is to account for the two reserved network IP Address - 0.xxx.xxx.xxx

(IP Address 0.0.0.0 is reserved for broadcasting requirements)

127.xxx.xxx.xxx

(IP Address 127.0.0.0 is reserved for loopback address used for software testing)

In all the classes, total numbers of Hosts that can be configured are 2 less.

This is to account for the two reserved IP address in which all the bits for Host ID are either zero or one.

When all Host ID bits are 0

(It represents the Network ID for the network)

When all Host ID bits are 1

(It represents the Broadcast Address)

8. Class B

Starts with "10"

Total Number of IP Addresses available in Class B = Numbers possible due to remaining available 30 bits = 2^{30}	Total number of Hosts that can be configured in Class B = Numbers possible due to available 16 bits in the Host Id -2 = $2^{16}-2$
32 bits	Range of 1 st Octet Minimum value of 1 st Octet = 10000000 = 128 Maximum value of 1 st Octet = 10111111 = 191 So, Range=[128,191]
Total number of networks available in Class B = Numbers possible due to remaining available 14 bits in the Net ID = 2^{14}	Class B is used by organizations requiring medium size networks like IRCTC, banks etc.

9. Class C

Starts with "110"

Total Number of IP Addresses available in Class C = Numbers possible due to remaining available 29 bits = 2^{29}	Total number of Hosts that can be configured in Class C = Numbers possible due to available 8 bits in the Host Id -2 = 2^8-2
--	--

32 bits	Range of 1 st Octet Minimum value of 1 st Octet = 11000000 = 192 Maximum value of 1 st Octet = 11011111 = 223 So, Range=[192, 223]
Total number of networks available in Class C = Numbers possible due to remaining available 21 bits in the Net ID =2 ²¹	Class C is used by organizations that require small to medium size networks like engineering colleges, small universities, small offices etc.

10. Class D

Starts with "1110"

Total Number of IP Addresses available in Class D = Numbers possible due to remaining available 28 bits =2 ²⁸	Range of 1 st Octet Minimum value of 1 st Octet = 11100000 = 224 Maximum value of 1 st Octet = 11101111 = 239 So, Range=[224, 239]
32 bits	Class D is reserved for multicasting. In multicasting, there is no need to extract host address from IP address because data is not destined for a particular host.

11. Class E

Starts with "1111"

Total Number of IP Addresses available in Class E = Numbers possible due to remaining available 28 bits =2 ²⁸	Range of 1 st Octet Minimum value of 1 st Octet = 11110000 = 240 Maximum value of 1 st Octet = 11111111 = 255 So, Range=[240, 255]
IP Address 32 bits	Class E is reserved for future or experimental purposes.

12. Keywords

Bluetooth: Bluetooth is a proprietary open wireless technology standard for exchanging data over short distances (using short wavelength radio transmissions in the ISM band from 2400-2480 MHz) from fixed and mobile devices, creating **personal** area networks (PANs) with high levels of security

Public IP Address: it is an address provided to a communication device by a internet service provider.

Dynamic IP addresses: these are the IP addresses which changed regularly and automatically. It is an IP address that have their own FTP service. It also provides the service of anonymous FTP sharing

Summary

The internet protocol is a part of the TCP IP protocol suite. The internet protocol is essential in assigning a unique IP address to a machine connected to the internet. Initially the IP addresses

came in the form of IPv4 which was a 32 bit logical address which was represented as four octets of decimal numbers separated by colons these decimal numbers could take values between 0 to 255. However with extensive usage the maximum limit of the IP addresses is about to be reached. So to fulfill the need of a broader range for identifying numerous communication devices connected to the internet IPv6 was devised. It is a 128 bit hexadecimal address which can support up to 340 undecillion addresses. Further to support more number of computers in a network with limited IP addresses the concept of subnetting can overcome this limitation.

Self Assessment

Fill in the blanks:

1. The IP addresses 10000001 00001011 00001011 11101111 in binary notation is equivalent to _____ dotted-decimal notation.
2. The IP addresses 11000001 10000011 00011011 11111111 in binary notation is equivalent to _____ dotted-decimal notation.
3. The IP addresses 11100111 11011011 10001011 01101111 in binary notation is equivalent to _____ dotted-decimal notation.
4. The IP addresses 11111001 10011011 11111011 00001111 in binary notation is equivalent to _____ dotted-decimal notation.
5. The IP addresses 111.56.45.78 in the dotted-decimal notation is equivalent to _____ in binary notation.
6. The IP addresses 221.34.7.82 in the dotted-decimal notation is equivalent to _____ in binary notation.
7. The IP addresses 241.8.56.12 in the dotted-decimal notation is equivalent to _____ in binary notation.
8. The IP addresses 75.45.34.78 in the dotted-decimal notation is equivalent to _____ in binary notation.

Multiple Choice Questions:

9. Choose the address of class D
 - A. Unicast
 - B. Reserved
 - C. Multicast
 - D. None of the given choices
10. Select the wrong class.
 - A. CLASS A = 1 to 126
 - B. CLASS C = 192 to 220
 - C. CLASS B = 128 to 191
 - D. CLASS D = 224 to 239
11. Which of the following is true in context to ARP?
 - A. ARP is used to find the MAC (Media Access Control) address of a device from its IP address.
 - B. ARP is used when a device wants to communicate with another device on a Local Area Network or Ethernet.
 - C. At DLL we need to know the MAC address of the machine by sending an ARP request packet
 - D. All the given choices
12. Which of the following is a type of Address Resolution Protocol
 - A. Gratuitous ARP

- B. Reverse ARP (RARP)
 - C. Inverse ARP
 - D. All the given choices
13. Which of the following is not true regarding Network Address Translation (NAT)?
- A. NAT allows multiple devices to access the Internet through a single public address.
 - B. The translation of private IP address to a public IP address is required.
 - C. One or more local IP address is translated into one or more Global IP address and vice versa to provide Internet access to the local hosts.
 - D. It does not perform the translation of port numbers
14. Which of the following is considered to be the address before translation?
- A. Inside Local
 - B. Inside Global
 - C. Outside Local
 - D. Outside Global
15. Which of the following is not an advantages of using NAT?
- A. Translation introduces switching path delays.
 - B. Conserves legally registered addresses.
 - C. Increases flexibility when connecting to the Internet.
 - D. Reduces address overlap occurrence.
16. If Direct Broadcast address is 201.15.16.31, which of the following can be subnet mask?
- A. 255.255.255.240
 - B. 255.255.255.248
 - C. 255.255.255.252
 - D. All the given choices
17. In class B if subnet mask is 255.192.0.0 Total Number of networks than can be joined is ____.
- A. 32
 - B. 64
 - C. 16
 - D. None of the given choices

Answer for self Assessment:

- | | |
|-------------------------------------|--|
| 1. 129.11.11.239 | 2. 193.131.27.255 |
| 3. 231.219.139.111 | 4. 249.155.251.15 |
| 5. 01101111001110000010110101001110 | 6. 11011101 00100010 00000111 01010010 |
| 7. 11110001000010000011100000001100 | 8. 01001011 00101101 00100010 01001110 |
| 9. C | 10. B |
| 11. D | 12. D |
| 13. D | |

14. A 15. A 16. D 17. B

Further Readings



Andrew S. Tanenbaum, *Computer Networks*, Prentice Hall.

Behrouz A. Forouzan and Sophia Chung Fegan, *Data Communications and Networking*, McGraw-Hill Companies.

J. D. Spragins, *Telecommunications Protocols and Design*, Addison-Wesley.

Rajneesh Agrawal and Bharat Bhushan Tiwari, *Computer Networks and Internet*, Vikas Publication



<https://www.geeksforgeeks.org/basics-computer-networking/>

Unit 09: Network Layer – Routing

CONTENTS

Objectives

Introduction

- 9.1 Network Layer Routing
- 9.2 Routing
- 9.3 Routing v/s Flooding:
- 9.4 Adaptive v/s Non-Adaptive Routing Algorithm
- 9.5 Unicast Routing
- 9.6 Broadcast Algorithms:
- 9.7 Spanning-Tree Creation
- 9.8 Multicasting
- 9.9 Multicasting - Applications
- 9.10 Multicast vs Multiple Unicast
- 9.11 Why Multicasting?
- 9.12 Why Group E-mail is Multiple Unicast?
- 9.13 Multicasting Challenges
- 9.14 Multicast Address
- 9.15 Delivery at Datalink Layer
- 9.16 Solution Characteristics

Summary

Keywords

Self Assessment

Answer for Self Assessment

Review Questions

Further Reading

Objectives

After this lecture, you would be able to

- understand the concept of Network Layer Routing
- learn the concept of Unicast routing and understand the major unicast routing protocols
- understand the role and classifications of various routing algorithms
- analyze the difference between adaptive and non-adaptive routing algorithms.
- understand the basic differences between unicast, multicast and broadcast routing
- learn the various algorithms available for the different routing schemes
- learn the routing techniques in Mobile Adhoc Networks
- understand the difference between reactive and proactive routing algorithms
- learn the concept of weighted graphs and understand the process of calculating the shortest path using the Dijkstra's shortest path algorithm.

Introduction

The network layer deals with forwarding packets from the source node to the destination node using different routes. Hence, the network layer transports traffic between devices that are not locally attached. In doing so, it controls the operation of the subnet, which involves routing of the packets from the source to destination. Routes are based on static or dynamic routing tables. The destination IP address is checked for packet received on a router interface. If the packet is not addressed for the router where it is received, the router will look up the destination network address in the routing table so that it may be routed accordingly. Therefore, the network layer must know about the topology of the communication subnet and choose appropriate paths through it. The routes are chosen in such a manner so that the network layer avoids overloading some of the communication lines while leaving others idle. The routing algorithm is part of the network layer. The routing algorithm enables the network layer to decide to which output line an incoming packet should be forwarded. Routing algorithms provides correctness, simplicity, robustness, stability, fairness and optimality. All these functions of network layer differ from the data link layer whose objective is to transmit the bits from one end of a wire to the other end. The Network layer is the lowest layer that deals with end-to-end transmission.

9.1 Network Layer Routing

When a device has multiple paths to reach a destination, it always selects one path by preferring it over others. This selection process is termed as Routing. Routing is done by special network devices called routers or it can be done by means of software processes. The software based routers have limited functionality and limited scope.

A router is always configured with some default route. A default route tells the router where to forward a packet if there is no route found for specific destination.

In case there are multiple path existing to reach the same destination, router can make decision based on the following information:

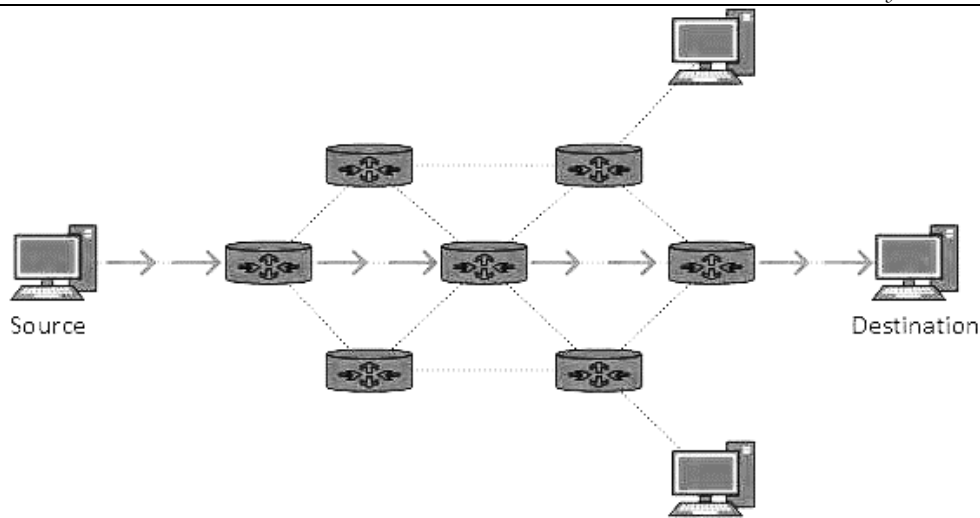
- Hop Count
- Bandwidth
- Metric
- Prefix-length
- Delay

Routes can be statically configured or dynamically learnt. One route can be configured to be preferred over others. The routing techniques can be broadly categorized into four categories:

- 1) **Unicast Routing**
- 2) **Broadcast routing**
- 3) **Multicast Routing**
- 4) **Any cast Routing**

1) Unicast Routing

Most of the traffic on the internet and intranets known as unicast data or unicast traffic is sent with specified destination. Routing unicast data over the internet is called unicast routing. It is the simplest form of routing because the destination is already known. Hence the router just has to look up the routing table and forward the packet to next hop.



Unicast means the transmission from a single sender to a single receiver. It is a point to point communication between sender and receiver.

There are various unicast protocols such as TCP, HTTP, etc.

- TCP is the most commonly used unicast protocol. It is a connection oriented protocol that rely on acknowledgement from the receiver side.
- HTTP stands for Hyper Text Transfer Protocol. It is an object oriented protocol for communication.

There are three major protocols for unicast routing:

1. Distance Vector Routing
2. Link State Routing
3. Path-Vector Routing

1. Distance Vector Routing

A distance-vector routing (DVR) protocol requires that a router inform its neighbors of topology changes periodically. Historically known as the old ARPANET routing algorithm (or known as Bellman-Ford algorithm).

Bellman Ford Basics

Each router maintains a Distance Vector table containing the distance between itself and ALL possible destination nodes. Distances, based on a chosen metric, are computed using information from the neighbors' distance vectors.

Information kept by DV router -

- Each router has an ID
- Associated with each link connected to a router, there is a link cost (static or dynamic).
- Intermediate hops

Distance Vector Table Initialization -

- Distance to itself = 0
- Distance to ALL other routers = infinity number.
- Distance Vector Algorithm
 1. A router transmits its distance vector to each of its neighbors in a routing packet.
 2. Each router receives and saves the most recently received distance vector from each of its neighbors.
 3. A router recalculates its distance vector when:
 - It receives a distance vector from a neighbor containing different information than before.

It discovers that a link to a neighbor has gone down.

The DV calculation is based on minimizing the cost to each destination

$D_x(y)$ = Estimate of least cost from x to y

$C(x, v)$ = Node x knows cost to each neighbor v

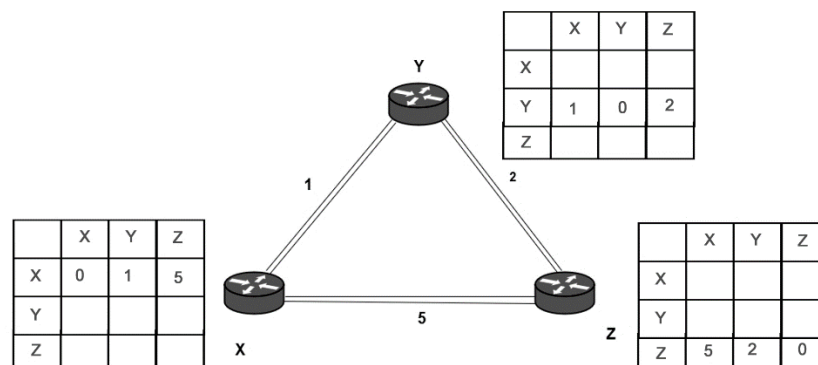
$D_x = [D_x(y): y \in N]$ = Node x maintains distance vector

Node x also maintains its neighbors' distance vectors

For each neighbor v, x maintains $D_v = [D_v(y): y \in N]$

- From time-to-time, each node sends its own distance vector estimate to neighbors.
- When a node x receives new DV estimate from any neighbor v, it saves v's distance vector and it updates its own DV using B-F equation:

$$D_x(y) = \min \{C(x, v) + D_v(y), D_x(y)\} \text{ for each node } y \in N$$

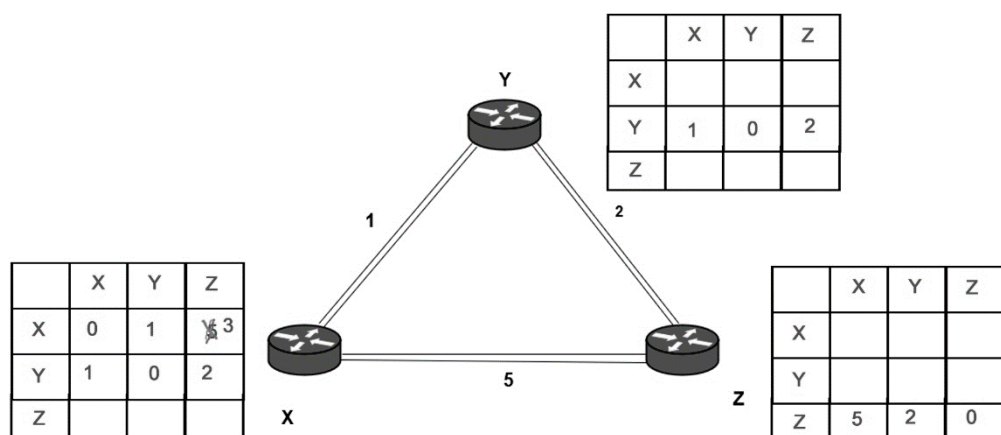


Consider 3-routers X, Y and Z as shown in figure. Each router has their routing table. Every routing table will contain distance to the destination nodes.

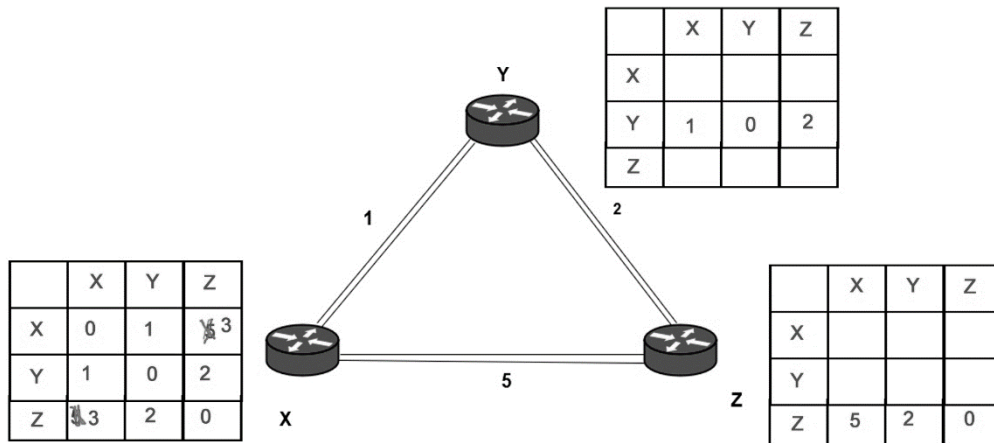
Consider router X, X will share its routing table to neighbors and neighbors will share their routing table to it. X and distance from node X to destination will be calculated using Bellman-Ford equation.

Bellman-Ford equation

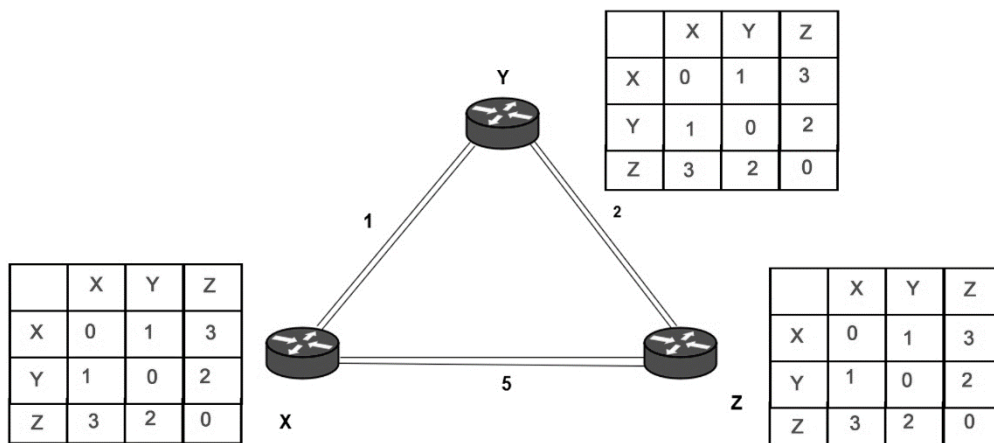
$$D_x(y) = \min \{C(x, v) + D_v(y)\} \text{ for each node } y \in N$$



As we can see that distance will be less going from X to Z when Y is intermediate node (hop) so it will be updated in routing table X.



Similarly for Z also



Finally the routing table for all –

Advantages of Distance Vector Routing

- It is simpler to configure and maintain than link state routing.

Disadvantages of Distance Vector Routing

- It is slower to converge than link state.
- It is at risk from the count-to-infinity problem.
- For larger networks, distance vector routing results in larger routing tables than link state since each router must know about all other routers. This can also lead to congestion on WAN links.
- It creates more traffic than link state since a hop count change must be propagated to all routers and processed on each router. Hop count updates take place on a periodic basis, even if there are no changes in the network topology, so bandwidth-wasting broadcasts still occur.

Note: Distance Vector routing uses UDP (User datagram protocol) for transportation.

2. Link State Routing

Link state routing is the second family of routing protocols. While distance vector routers use a distributed algorithm to compute their routing tables, link-state routing uses link-state routers to exchange messages that allow each router to learn the entire network topology. Based on this learned topology, each router is then able to compute its routing table by using a shortest path computation.

Features of Link State Routing Protocols

- Link state packet – A small packet that contains routing information.
- Link state database – A collection information gathered from link state packet.
- Shortest path first algorithm (Dijkstra algorithm) – A calculation performed on the database results into shortest path
- Routing table – A list of known paths and interfaces.

Calculation of shortest path –

To find shortest path, each node need to run the famous Dijkstra algorithm. This famous algorithm uses some steps.

These are:

Step-1: The node is taken and chosen as a root node of the tree, this creates the tree with a single node, and now set the total cost of each node to some value based on the information in Link State Database

Step-2: Now the node selects one node, among all the nodes not in the tree like structure, which is nearest to the root, and adds this to the tree. The shape of the tree gets changed.

Step-3: After this node is added to the tree, the cost of all the nodes not in the tree needs to be updated because the paths may have been changed.

Step-4: The node repeats the Step 2 and 3. Until all the nodes are added in the tree

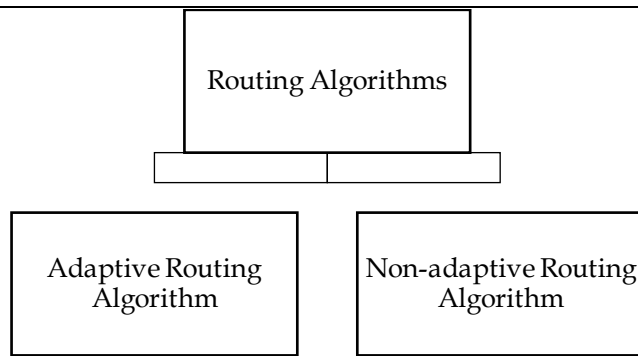
9.2 Routing

It is the process of establishing the routes that data packets must follow to reach the destination. In this process, a routing table is created which contains information regarding routes which data packets follow. Various routing algorithm are used for the purpose of deciding which route an incoming data packet needs to be transmitted on to reach destination efficiently.

- Routing Algorithm
- In order to transfer the packets from source to the destination, the network layer must determine the best route through which packets can be transmitted.
- Whether the network layer provides datagram service or virtual circuit service, the main job of the network layer is to provide the best route. The routing protocol provides this job.
- Routing Algorithm
- The routing protocol is a routing algorithm that provides the best path from the source to the destination. The best path is the path that has the "least-cost path" from source to the destination.
- Routing is the process of forwarding the packets from source to the destination but the best route to send the packets is determined by the routing algorithm.
- Classification of Routing Algorithms

The Routing algorithm is divided into two categories:

- Adaptive Routing algorithm
- Non-adaptive Routing algorithm



1) Adaptive Routing Algorithm

- This algorithm makes the routing decisions based on the topology and network traffic.
- Also known as dynamic routing, these make use of dynamic information such as current topology, load, delay, etc. to select routes.
- The changes in routing decisions are reflected in the topology as well as traffic of the network.
- The main optimizing parameters related to this algorithm are hop count, distance and estimated transit time.

Adaptive Routing Algorithm Types

An adaptive routing algorithm can be classified into three categories:

- Centralized algorithm
- Isolation algorithm
- Distributed algorithm

a) Centralized Algorithm

It is also known as global routing algorithm. It computes the least-cost path between source and destination by using complete and global knowledge about the network. It takes the connectivity between the nodes and link cost as input, and this information is obtained before actually performing any calculation. Link state algorithm is referred to as a centralized algorithm since it is aware of the cost of each link in the network. In this method, a centralized node has entire information about the network and makes all the routing decisions.

Advantage:

- Only one node is required to keep the information of entire network

Disadvantage:

- If central node goes down the entire network is done.
- Link state algorithm is referred to as a centralized algorithm since it is aware of the cost of each link in the network.

b) Isolation Algorithm

It is an algorithm that obtains the routing information by using local information rather than gathering information from other nodes and makes its routing decisions. The sending nodes don't have information about status of particular link.

Disadvantage:

- Packet may be sent through a congested network which may result in delay.



Hot potato routing, backward learning.

c) Distributed Algorithm

It is also known as decentralized algorithm as it computes the least-cost path between source and destination in an iterative and distributed manner. No node has the knowledge about the cost of all the network links. In the beginning, a node contains the information only about its own directly attached links and through an iterative process of calculation computes the least-cost path to the destination. In this method, the node receives information from its neighbors and then takes the decision about routing the packets. A Distance Vector Algorithm is a decentralized algorithm as it never knows the complete path from source to the destination; instead it knows the direction through which the packet is to be forwarded along with the least cost path.

Disadvantage:

- The packet may be delayed if there is change in between interval in which it receives information and sends packet.

2) Non-Adaptive Routing Algorithm

These are the algorithms which do not change their routing decisions once they have been selected. This is also known as static routing as route to be taken is computed in advance and downloaded to routers when router is booted.

- Non Adaptive routing algorithm is also known as a static routing algorithm.
- When booting up the network, the routing information stores to the routers.
- Non Adaptive routing algorithms do not take the routing decision based on the network topology or network traffic.

Non-Adaptive Routing Algorithm Types

The Non-Adaptive Routing algorithm is of two types:

1. Flooding.
2. Random walks.

1. Flooding

In case of flooding, every incoming packet is sent on every outgoing line except from which it arrived. One problem with this is that packets may go in loop and as a result of which a node may receive duplicate packets. These problems can be overcome with the help of sequence numbers, hop count and spanning tree. In case of flooding, every incoming packet is sent on every outgoing line except from which it arrived. One problem with this is that packets may go in loop and as a result of which a node may receive duplicate packets. These problems can be overcome with the help of sequence numbers, hop count and spanning tree. In case of flooding, every incoming packet is sent on every outgoing line except from which it arrived. One problem with this is that packets may go in loop and as a result of which a node may receive duplicate packets. These problems can be overcome with the help of sequence numbers, hop count and spanning tree.

Disadvantage:

- A node may contain several copies of a particular packet.
- In case of random walks, a packet sent by the node to one of its neighbors randomly.
- In this method, packets are sent host by host or node by node to one of its neighbors randomly.
- This is highly robust method which is usually implemented by sending packets onto the link which is least queued.

Advantage

The advantage of using random walks is that it uses the alternative routes very efficiently.

9.3 Routing v/s Flooding:

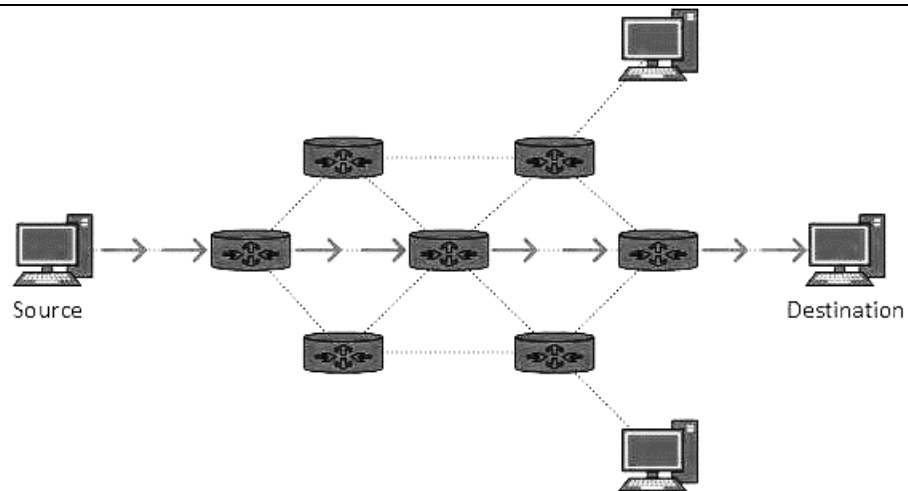
Sr.	Routing	Flooding
1	Routing Table is required.	No routing table is required.
2	May give shortest path.	Always gives shortest path.
3	Less reliable.	More reliable.
4	Traffic is less.	Traffic is high.
5	No duplicate packets.	Duplicate packets are there.

9.4 Adaptive v/s Non-Adaptive Routing Algorithm

Basis of Comparison	Adaptive Routing Algorithm	Non-Adaptive Routing Algorithm
Define	Adaptive Routing algorithm is an algorithm that constructs the routing table based on the network conditions.	The Non-Adaptive Routing algorithm is an algorithm that constructs the static table to determine which node to send the packet.
Usage	Adaptive routing algorithm is used by dynamic routing...	The Non-Adaptive Routing algorithm is used by static routing.
Routing decision	Routing decisions are made based on topology and network traffic.	Routing decisions are the static tables.
Categorization	Traffic is less.	Traffic is high.
Complexity	No duplicate packets.	Duplicate packets are there.

9.5 Unicast Routing

- In unicast routing, there is one source and one destination node i.e. point-to-point communication.
- The relationship between the source and the destination network is one to one.
- Each router in the path tries to forward the packet to one and only one of its interfaces.
- Routing unicast data over the internet where unicast traffic is sent with specified destination which is already known to the router is called unicast routing.
- Hence the router just has to look up the routing table and forward the packet to next hop.



2) Broadcast Routing

- By default, the broadcast packets are not routed and forwarded by the routers on any network.
- Routers create broadcast domains. But it can be configured to forward broadcasts in some special cases.
- A broadcast message is destined to all network devices.
- In broadcast routing, the network layer provides a service of delivering a packet sent from a source node to all other nodes in the network.
- In broadcast routing, packets are sent to all nodes even if they do not want it.

Broadcast Routing can be done in Two Ways:

Method 1:

- A router creates a data packet and then sends it to each host one by one.
- In this case, the router creates multiple copies of single data packet with different destination addresses.
- All packets are sent as unicast but because they are sent to all, it simulates as if router is broadcasting.
- This method consumes lots of bandwidth and router must know the destination address of each node.

Method 2:

- When router receives a packet to be broadcasted, it simply floods those packets out of all interfaces.
- All routers are configured in the same way.
- This method is easy on router's CPU but may cause the problem of duplicate packets received from peer routers.
- Reverse path forwarding is a technique, in which router knows in advance about its predecessor from where it should receive broadcast.
- This technique is used to detect and discard duplicates.
- In multicast routing, a single source node can send a copy of a packet to a subset of the other network nodes.

Most straightforward way: N-way-unicast

- No new network-layer routing protocol, packet-duplication, or forwarding functionality is needed.

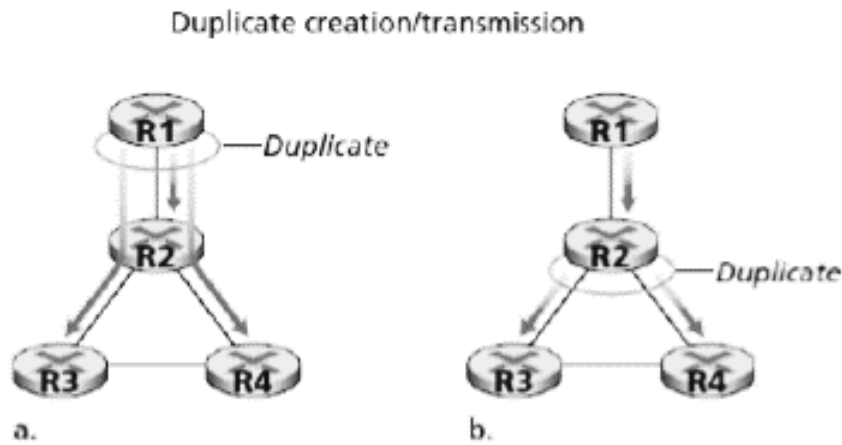


Fig: Source-duplication versus In-network duplication

Drawbacks:

- Inefficiency: As it would be more efficient for the network nodes themselves (rather than just the source node) to create duplicate copies of a packet
- Unrealistic assumption: An implicit assumption of N-way-unicast is that broadcast recipients, and their addresses, are known to the sender.
- More overhead: it would be unwise (at best!) to rely on the unicast routing infrastructure to achieve broadcast.

9.6 Broadcast Algorithms:

1. Uncontrolled Flooding
2. Controlled Flooding
3. Spanning Tree Broadcast etc.

1. Uncontrolled Flooding

- Most obvious technique for achieving broadcast is a flooding.
- Source node sends a copy of the packet to all of its neighbors
- When a node receives a broadcast packet, it duplicates the packet and forwards it to all of its neighbors (except the neighbor from which it received the packet).
- This scheme will eventually deliver a copy of the broadcast packet to all nodes if they are connected

Disadvantages:

- If the graph has cycles, then one or more copies of each broadcast packet will cycle indefinitely
- When a node is connected to more than two other nodes, then it could result in broadcast storm (resulting from the endless multiplication of broadcast packets)

2. Controlled Flooding

- It is a key to avoiding a broadcast storm
 - For a node to judiciously choose when to flood and when not to flood a packet
 - i.e. controlled way of flooding
- A source node puts its address as well as a broadcast sequence number into a broadcast packet
- Each node maintains a list of the source address and sequence number of each broadcast packet it has already received, duplicated, and forwarded

When a node receives a broadcast packet, it first checks in this list.

- If found, then dropped the packet

- If not found, then the packet is duplicated and forwarded to all the node's neighbors (except the node from which the packet has just been received)

Reverse Path Forwarding (RPF) / Reverse Path Broadcast (RPB)

When a router receives a broadcast packet with a given source address,

- It transmits the packet on all of its outgoing links (except the one on which it was received)

Only if the packet arrived on the link that is on its own shortest unicast path back to the source.

Otherwise, it simply discards the incoming packet.

- RPF does not use unicast routing to actually deliver a packet to a destination, nor does it require that a router know the complete shortest path from itself to the source.
- RPF need only know the next neighbor on its unicast shortest path to the sender

3. Spanning Tree Broadcast

While sequence-number-controlled flooding and RPF avoid broadcast storms,

- they do not completely avoid the transmission of redundant broadcast packets

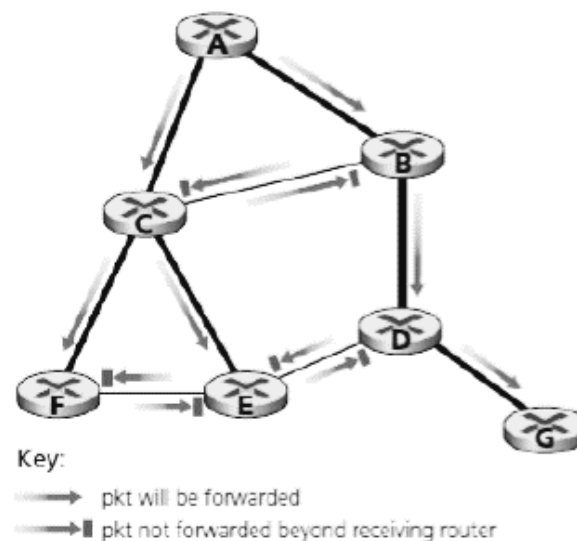


Fig. Reverse Path Forwarding

In this figure, nodes B, C, D, E, and F receive either one or two redundant packets.

Solution: spanning tree – a tree that contains each and every node in a graph.

So, first construct a spanning tree.

- When a source node wants to send a broadcast packet,
- It sends the packet out on all of the incident links that belong to the spanning tree.

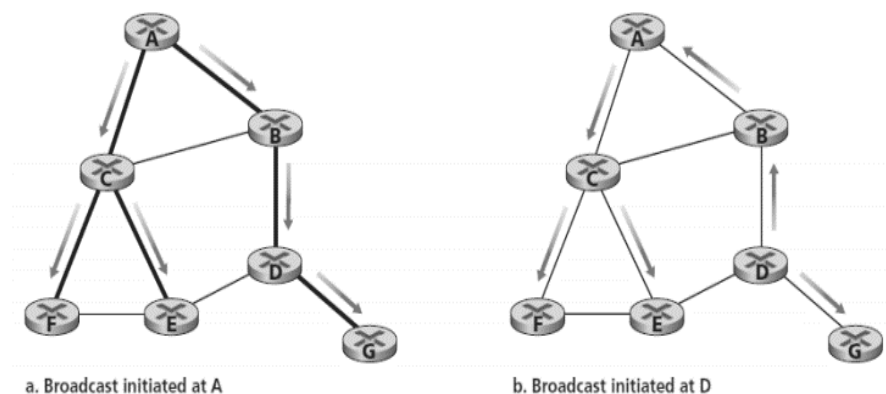


Fig: Broadcast along a Spanning Tree

- Not only does spanning tree eliminate redundant broadcast packets, but once in place, the spanning tree can be used by any node to begin a broadcast
- In this algorithm, a node need not be aware of the entire tree; it simply needs to know which of its neighbors in G are spanning-tree neighbors

9.7 Spanning-Tree Creation

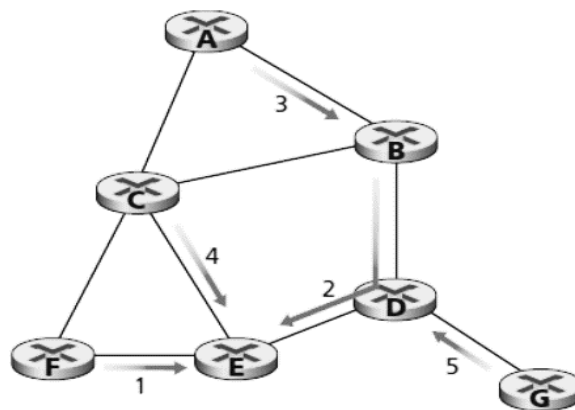
- The main complexity associated with the spanning-tree based broadcast approach is the creation and maintenance of the spanning tree.
- One simple algorithm is center-based approach

At first a center node or a core is defined

Each nodes then unicast tree-join messages addressed to the center node

A tree-join message is forwarded using unicast routing toward the center

- Until it either arrives at a node that already belongs to the spanning tree or arrives at the center.



a. Stepwise construction of spanning tree

Considering node E as core

9.8 Multicasting

- There is one source and a group of destinations, but not all.
- The relationship is one too many.
- The source address is a unicast address,

but the destination address is a group address,

In which there is at least one member of the group that is interested in receiving the multicast datagram.

9.9 Multicasting - Applications

Few Applications of Multicasting:

- bulk data transfer to a group
- streaming continuous media
- shared data applications (e.g. teleconferencing)
- Web cache updating

- interactive gaming

Multicast vs Multiple Unicast

Multicasting

- Starts with a single packet from source that is duplicated by the routers.
- The destination address in each packet is the same for all duplicates.
- Only a single copy of the packet travels between any two routers.
- IP Multicast uses UDP for communication, therefore it is unreliable.

Multiple Unicasting

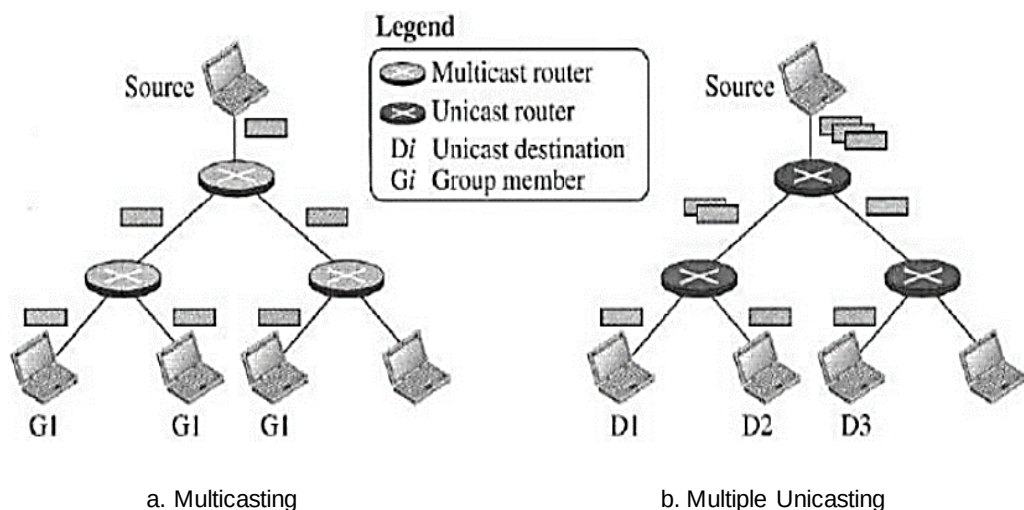
- several packets start from the source.
- If there are three destinations, the source sends three packets, each with a different unicast destination address.

There may be multiple copies traveling between two routers.

Group Email: When a person sends an e-mail message to a group of people, this is multiple Unicasting.

Teleconferencing: A group of workstations form a multicast group such that a transmission from any member is received by all other group members.

9.10 Multicast vs Multiple Unicast



9.11 Why Multicasting?

Two main reasons:

- Multicasting requires less bandwidth than multiple Unicasting.
- In multiple Unicasting, the packets are created by the source with a relative delay between packets.
- In multicasting, there is no delay because only one packet is created by the source.

9.12 Why Group E-mail is Multiple Unicast?

Multicast involves a subscription from the receiver's side,

- But, multiple unicast is a decision from the sender's side.
- Usually, sender manage the group of multiple unicast,
- But, a receiver is associated with a multicast group.

9.13 Multicasting Challenges

Two important problems:

1. how to identify the receivers of a multicast packet
2. how to address a packet sent to these receivers

Solution:

- a multicast packet is addressed using address indirection

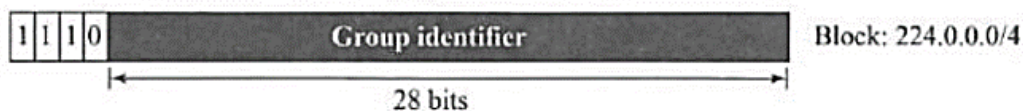
i.e., a single identifier is used for the group of receivers

- The group of receivers associated with such address is referred to as a multicast group.

IGMP is used to create and maintain multicast groups

9.14 Multicast Address

- In IP datagram, we can only write one destination address.
- So, we need multicast address for sending the datagram to many destinations.
- A multicast address is an identifier for a group.
- If a new group is formed with some active members, an authority can assign an unused multicast address to this group to uniquely define it
- A router / a destination host needs to distinguish between a unicast and a multicast datagram.
- IPv4 assigns a block of addresses for this purpose
- In classful addressing, all of class D was composed of these addresses;
- In classless addressing, it is referred to as the block 224.0.0.0/4 (i.e., 224.0.0.0 - 239.255.255.255).



9.15 Delivery at Data link Layer

- In multicasting, the delivery at the Internet level is done using multicast IP addresses
- But, data-link layer multicast addresses are also needed to deliver a multicast packet encapsulated in a frame.
- Address Resolution Protocol (ARP) cannot help in finding multicast MAC address

Solution for two scenarios:

- Network with Multicast Support
- Network with No Multicast Support

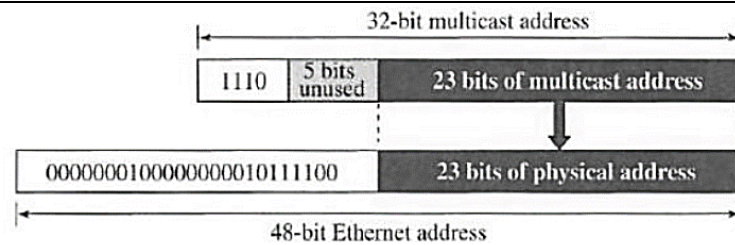
Case 1: Network with Multicast Support

Most LANs (e.g. Ethernet) support physical multicast addressing.

If the first 25 bits in an Ethernet address are

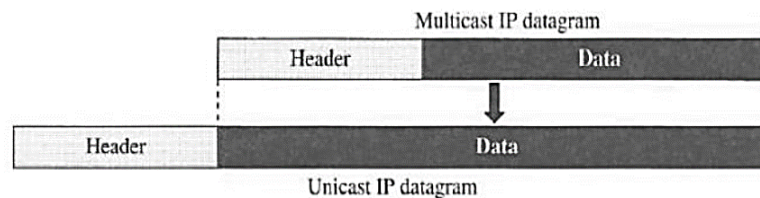
0000 0001 0000 0000 0101 1110 0

This identifies a physical multicast address for the TCP/IP protocol.



Case 2. Network with No Multicast Support

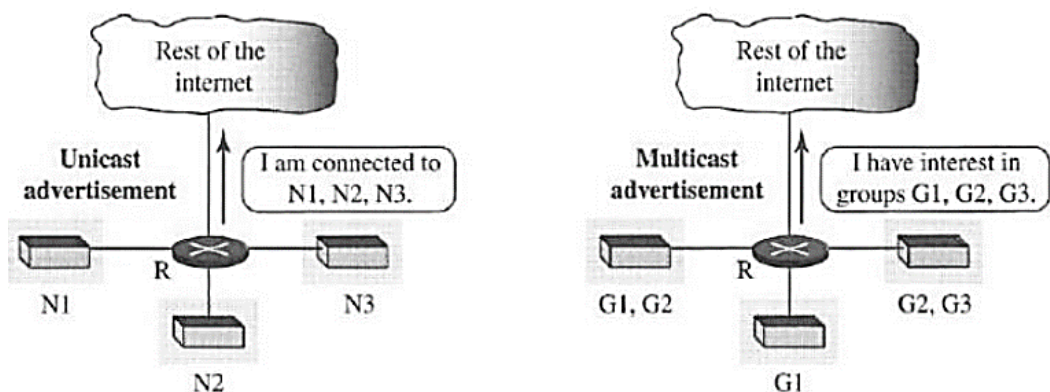
Most WANs do not support physical multicast addressing. To send a multicast packet through these networks, a tunneling is used. In tunneling, the multicast packet is encapsulated in a unicast packet and sent.



Collecting Information about Groups

The creation of forwarding tables in both unicast and multicast routing involves two steps:

- A router needs to know to which destinations it is connected.
- Each router needs to propagate information obtained in the first step to all other routers so that each router knows to which destination each other router is connected



- In unicast routing, the collection of the information in the first step is automatic. Each router knows to which network it is connected, and the prefix of the network (in CIDR) is what a router needs.
- In multicast routing, the collection of information in the first step is not automatic because, a router does not know which host in the attached network is a member of a particular group; Membership in the group does not have any relation to the prefix associated with the network. the membership is not a fixed attribute of a host; A host may join some new groups and leave some others even in a short period of time.

For Unicasting, the router needs no help to collect; but for multicasting, it needs the help of another protocol namely Internet Group Management Protocol (IGMP)

IGMP: Internet Group Management Protocol

IGMP messages, like ICMP messages, are carried (encapsulated) within an IP datagram. IGMP uses three messages: Query, Report, and Leave

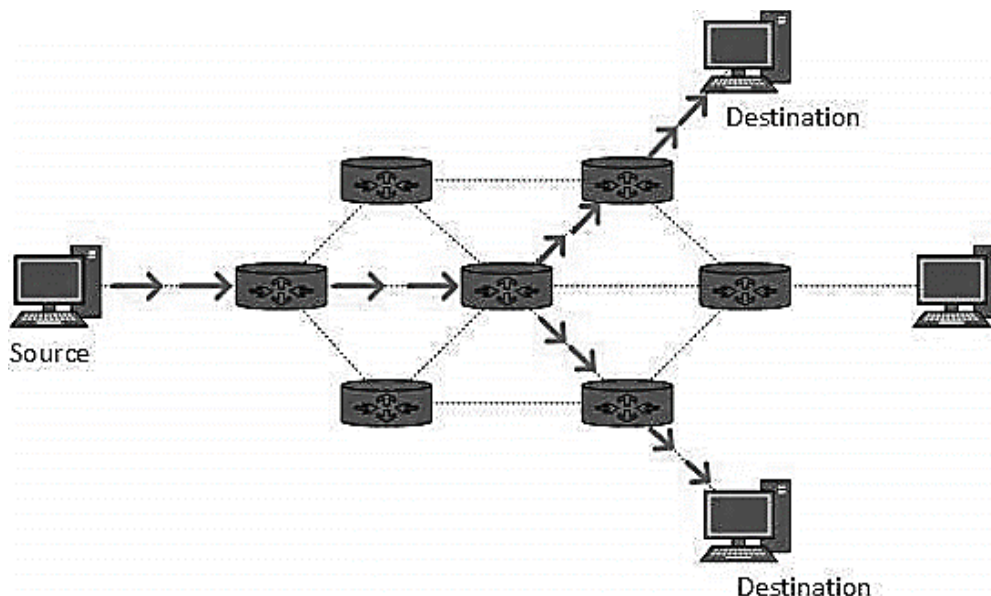
- A **query** message is periodically sent by a router to all hosts attached to it to ask them to report their interests about membership in groups.
- A **report** message is sent by a host as a response to a query message. After a router has collected membership information from the hosts and other routers at its own level in the tree, it can propagate the information to the router located in a higher level of the tree.
- **Leave** group message is used to inform its leaving. This message is optional.

Multicast Forwarding

- a router needs to make a decision to forward a multicast packet
- In unicast communication, the destination address of the packet defines one single destination. So, forwarded through one interface.
- In multicast communication, the destination of the packet defines one group, but that group may have more than one member in the internet. So, forwarded through many interfaces.
- Forwarding decisions in unicast communication depend only on the destination address of the packet.
- Forwarding decisions in multicast communication depend on both the destination and the source address of the packet.

Multicast Routing

- In multicast routing, a single source node can send a copy of a packet to a subset of the other network nodes.
- Multicast routing is special case of broadcast routing with significance difference and challenges.
- But in Multicast routing, the data is sent to only nodes which wants to receive the packets.
- The router must know that there are nodes, which wish to receive multicast packets (or stream) then only it should forward.
- Multicast routing works spanning tree protocol to avoid looping.
- Multicast routing also uses reverse path Forwarding technique, to detect and discard duplicates and loops.



Any cast Routing

- Any cast packet forwarding is a mechanism where multiple hosts can have same logical address.
- When a packet destined to this logical address is received, it is sent to the host which is nearest in routing topology.

Mobile Ad-Hoc Networks

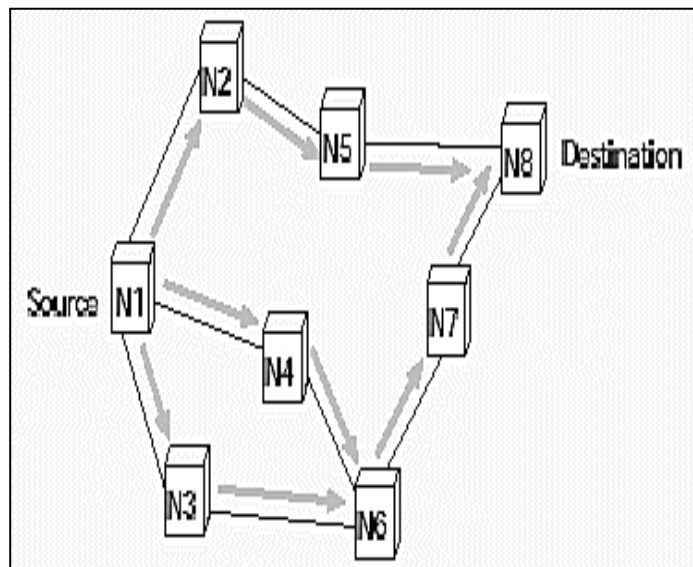
- Mobile Ad hoc Networks (MANET), are complex and distributed networks that are dynamic, infrastructure less and multi-hop in nature.
- The communication of a node can be either direct or through intermediate nodes without a fixed and dedicated infrastructure.
- Hence it is necessary to design an efficient routing protocol for ad hoc network which can address the issues of MANET efficiently.
- In ad hoc, routing algorithms are classified into nine categories.

The nine categories of ad-hoc routing algorithms are:

1. Source-initiated (reactive),
2. Table-driven (proactive),
3. Hybrid, hierarchical,
4. Multipath,
5. Multicast,
6. Location-aware,
7. Geographical-multicast and power-aware.

Routing in Mobile Ad-Hoc Networks

It determines and manages the routes between nodes in the Adhoc network.

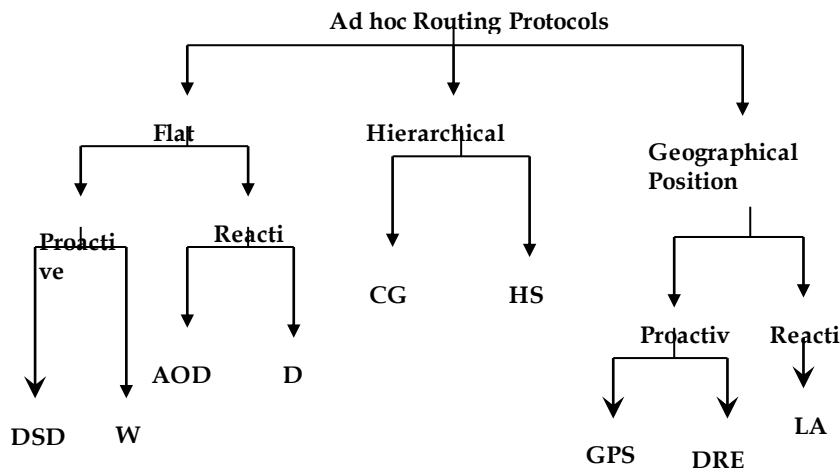
**9.16 Solution Characteristics**

- Different metrics to optimize
- Should address frequent changes in :
 - Topology
 - Traffic Patterns, etc.
- Handle any system characteristics such as unidirectional links.
- Should address trade-off between latency in determining a route and the overhead of route management (control traffic and state).

Routing Metrics

- Route length (#hops)
- Signal stability (SSR)
- Freshness of route (AODV)
- Association stability (ABR)

Classification



Adhoc Routing Protocols

1. Flat Routing Protocols

- Proactive Routing Protocols
 - DSDV Destination sequenced distance vector routing
 - WRP Wireless routing protocol
- Reactive Routing Protocols
 - AODV Adhocon-demand distance vector
 - DSR Dynamic source routing

2. Hierarchical Routing Protocols

- CGSR Cluster head gateway switch routing protocols
- HSR Hierarchical state routing

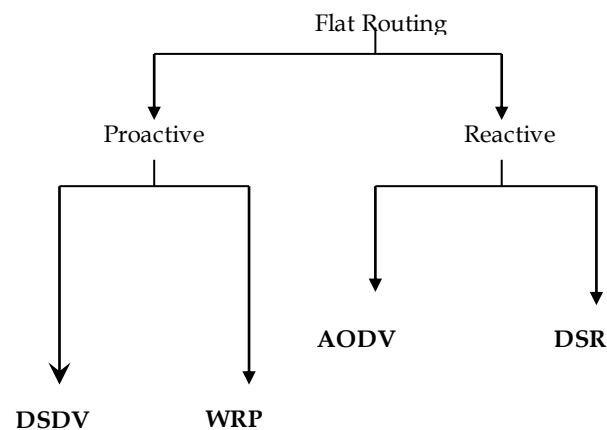
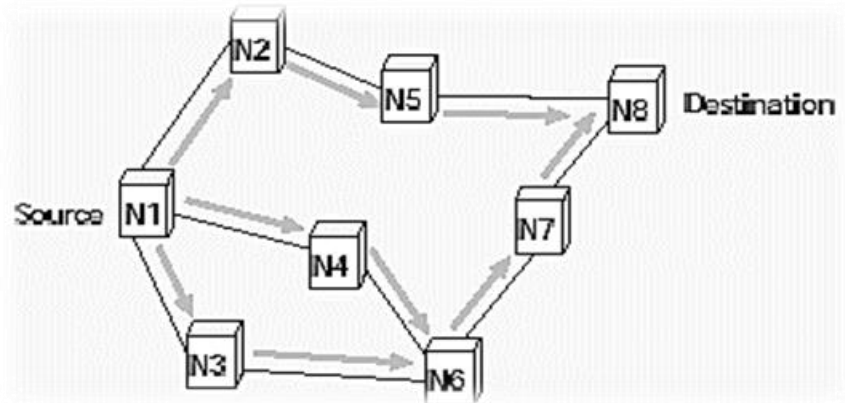
3. GPAR Geographical Position Augmented Routing

- Proactive
 - GPSR Greedy perimeter stateless routing
 - Dream Distance routing effect algorithm
- Reactive
 - LAR Location aided routing (LAR)

1. Flat Routing

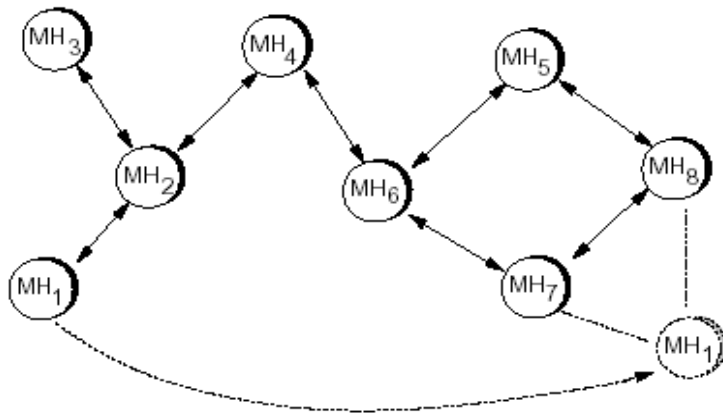
- Flat routing protocol is a network communication protocol implemented by routers in which all routers are each other's peers.
- Flat routing protocol distributes routing information to routers that are connected to each other without any organization or segmentation structure between them.
- Flat routing protocols are primarily those that don't work under a predefined network layout and perimeter.

- They enable the delivery of packets among routers through any available path without considering network hierarchy, distribution and composition.
- Flat routing protocols assume a flat topology
- All nodes are equally involved in routing.
- Flat routing protocol is implemented in flat networks where each router node routinely collects and distributes routing information with its neighboring routers.
- The entire participating node addressed by flat routing protocol performs an equal role in the overall routing mechanism.
- Routing Information Protocol, Interior Gateway Routing Protocol and Enhanced Interior Gateway Routing Protocol are popular examples of flat routing protocols.



1.1 Proactive Routing Protocols

- Maintain data structures which indicate the path to be taken for each destination node (also called table-driven protocols).
- Keep the data structures up-to-date by periodic exchange of information.



Destination	Metric	Next Hop
MH1	2	MH2
MH2	1	MH2
MH3	2	MH2
MH4	0	-
MH5	2	MH6
....		

Fig. Table for node MH4

In proactive (Table Driven) routing, each node has one or more tables that contain the latest information of the routes to any node in the network.

Out of the many different types the main prominent types of proactive routing are:

1. DSDV **Destination Sequenced Distance Vector Routing**
2. WRP **Wireless Routing Protocol**

1.1.1 Destination Sequenced Distance Vector Routing

Destination Sequenced Distance Vector Routing (DSDV): is mainly based on the Bellman Ford routing mechanism.

It is a table-driven routing protocol which was developed by Perkins and Bhagwat

1.1.2 Wireless Routing Protocol

The wireless routing protocol (WRP) was designed by Murthy and Garcia-Luna.

It uses the properties of the distributed Bellman-Ford algorithm.

In this algorithm, the route is chosen by selecting a neighbor node that would minimize the path cost.

1.2 Reactive Routing Protocols

In reactive routing protocols, a node initiates a route discovery, only when it wants to send packet to its destination.

They do not maintain or constantly update their route tables with the latest route topology.

Therefore, the communication overhead is reduced but the delay is increased due the on-demand route establishment process.

1.2 Reactive Routing Protocols

Determine a route to a destination only when a request for transmission to that node arrives (also called on-command or source-initiated protocols).

Usually cache newly found routes to minimize the latency involved in route discovery.

1.2 Reactive Routing Protocols

Involve three phases:

- Route discovery
- Route maintenance
- Route erasure

1.2 Reactive Routing

Out of the many different types the main prominent types of proactive routing are:

1. AODV **Destination Sequenced Distance Vector Routing**
2. DSR **Dynamic Source Routing**

1.2.1 Adhoc-on-demand Distance Vector
Routing

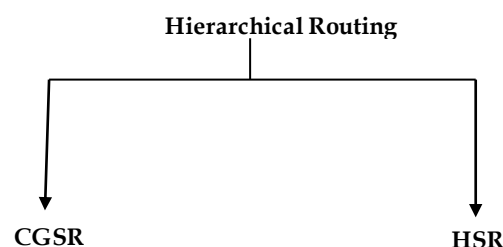
It was proposed by Perkins to provide loop- free routes even under the condition of repairing the failure routes.

The Time to Live (TTL), prevents the unnecessary forwarding of packets by a node hence reduces control overhead. Since, the performance depends on the bandwidth and end-end delay, so the route cache mechanism is not implemented in this protocol.

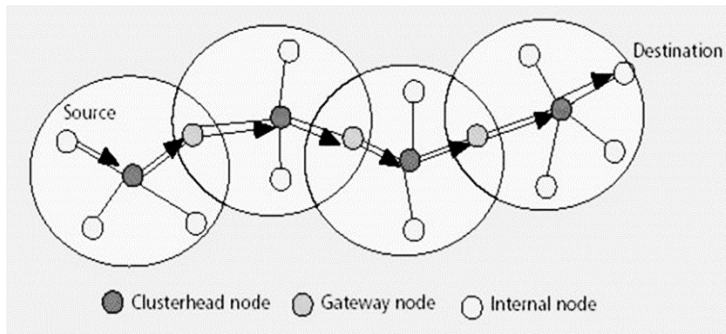
1.2.2 Dynamic Source Routing

- It is a primary on-demand routing protocol proposed by Johnson et al.
- The network bandwidth overhead is reduced by transmitting the routing message on-demand and battery power is harvested on the nodes since each of the nodes has to transmit the control packets whenever needed.
- It is a most widely known protocol that relays on source routing mechanism.

Hierarchical Routing



Hierarchical routing protocols choose some nodes which act as the backbone of the MANET.



2.1 Hierarchical Routing Protocols

- These protocols apply clustering techniques to build a hierarchy of nodes.
- Nodes are organized into groups called zones (or) clusters.
- Each cluster consists of one or more clusters and gateways.
- Hierarchical routing protocols are developed with an ability to address scalability issues in ad hoc network environment and to minimize excessive overhead.
- This on the other side increases the tediousness of the routing techniques used by these protocols.
- They can broadly be categorized in two types:
 1. Cluster Head Gateway Switch Routing Protocols (CGSR)
 2. Hierarchical state routing (HSR)

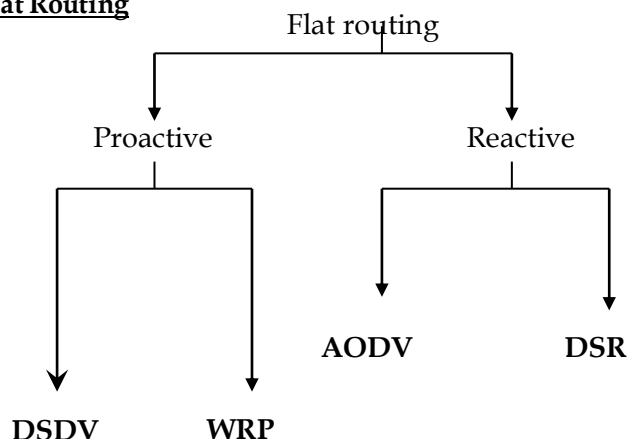
2.1.1 Cluster Head Gateway Switch Routing Protocols

- It employs a hierarchical network topology.
- It is based on a distributed algorithm namely least cluster change (LCC).
- Cluster head is elected by using least cluster change.
- Least Cluster Change algorithm is considered to be stable algorithm for cluster head election.
- Clustering enables an effective way for channel allocation.

2.1.2 Hierarchical Routing Protocols (HSR)

- It is a dynamic, distributed multilevel cluster based hierarchical protocol.
- In HSR clustering schema play a vital role.
- The primary objective of clustering is to have the efficient utilization of radio channel resource and the reduction of routing overhead, thus the network performance can be enhanced.

Flat Routing



Geographical Position Augmented Routing

- The geographical information about a node is collected by another node by using GPS mechanism.
- Location-aware routing protocols are efficiently supports to improve the scalability of the ad hoc network.
- Uses geographical information (say from GPS) in routing.

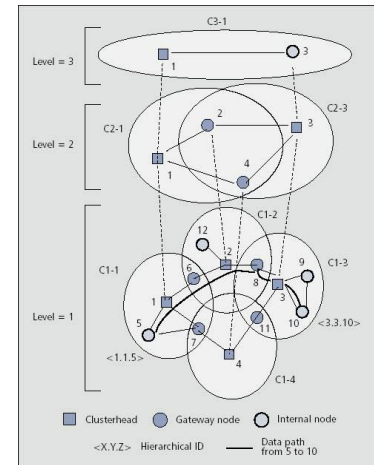
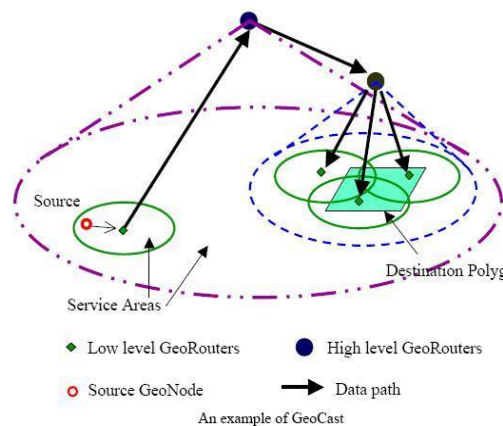


Figure 4. HSR: an example of multilevel clustering.

It can further be classified into two categories namely:

1. Proactive Routing
2. Reactive Routing

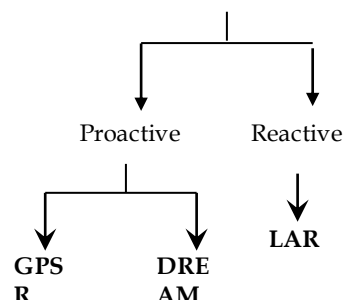
1 Proactive Routing Protocols

- Maintain data structures which indicate the path to be taken for each destination node (also called table-driven protocols).
- Keep the data structures up-to-date by periodic exchange of information.

2 Reactive Routing Protocols

- In reactive routing protocols, a node initiates a route discovery, only when it wants to send packet to its destination.
- They do not maintain or constantly update their route tables with the latest route topology.
- Therefore, the communication overhead is reduced but the delay is increased due the on-demand route establishment process.

Geographical Position Augmented Routing



3.2.1 Location Aided Routing Protocols

- It is based on directed flooding strategies.

- Two different LAR schemes were proposed by Ko and Vaidya to determine whether a node is within the request zone.

Summary

- The main role of the network layer is to accept packets from a source and deliver them to a destination machine. The network layer provides services that should be independent of the router technology. It shields the transport layer from the router network details and facilitates, network addressing to be consistent across networks.
- Services of the network layer are available in connection oriented and connection less modes. Connection-oriented services are useful only when the user wants to send a constant stream of data down the line.
- The routing algorithms that require selecting a path or route from many possible routes in the network are part of the router software. They are of two basic types namely non adaptive or static and dynamic or adaptive. Selection of routing algorithms depends on the minimum mean delay for the packets and number of hop before reaching to the destination machine.
- Link State Routing attempts to discover its neighbor and learn their network addresses and enable the router to choose a shortest path. The hierarchical routing uses multiple groups to route the packets. Broadcast and multicast routings are used to forward a single packet to several recipients depending on whether they belong to broadcast or multicast group.
- The shortest path to each destination within the network is found by traversing the tree and the most common shortest path first algorithm is the Dijkstra algorithm.
- The distance vector algorithms is used to determine which path is the best path to each destination based on the advertised details about the path and distance for each destination which is maintained in a local database.

Keywords

Adaptive Algorithms: They are capable of changing their routing decisions to reflect changes in the topology and the traffic and automatically update routing information when changes are made to the network configuration.

Distance Vector Routing: It maintains a routing table and exchanges its routing table with each of its neighbors so that their routing tables get updated.

Flow-based Routing: It considers the topology as well as the load.

Hierarchical Routing: It uses intra-domain routing and inter-domain routing.

Link State Routing: It enables each router in the network learns the network topology to creates a routing table based on this topology.

Multicast: It is used for one or more network interfaces located on various subnets. It allows one-to-many communication.

Multicast Routing: Refers to sending information to well-defined groups that have large members but small compared to the network as a whole.

Non-adaptive Algorithms: They are independent of the volume of the current traffic and to apology and decide the route to which a datagram is to send off-line.

Optimality Principle: This defines the optimal path.

Routing Algorithms: They are software part of the router and decide which output line an incoming packet should be transmitted on.

Self Assessment

1. A one-to-all communication between one source and all hosts on a network is classified as a _____ communication.
 - A. Broadcast
 - B. Unicast
2. A one-to-many communication between one source and a specific group of hosts is classified as a _____ communication.

- A. Multicast
- B. Broadcast

3. Which of the following is not true regarding Network Layer Routing?

- A. The software-based routers have limited functionality and limited scope.
- B. A default route tells the router where to forward a packet if there is no route found for specific destination.
- C. A Router cannot be configured to be preferred over others.
- D. Routers can be dynamically learnt

4. Which of the following are the major protocols of unicast routing?

- A: Distance Vector Routing
- B: Link State Routing
- C: Path-Vector Routing
- A. A and B
- B. A and C
- C. B and C
- D. A, B and C

5. Which of the following is not true regarding the Distance Vector Routing Protocol?

- A. It requires that a router inform its neighbors of topology changes periodically.
- B. It is also known as the old ARPANET routing algorithm
- C. It is also known as the Bellman-Ford algorithm
- D. Each router doesn't require to maintain a Distance Vector routing table

6. What is not true regarding the adaptive routing algorithms?

- A. It makes the routing decisions based on the topology and network traffic.
- B. It makes use of dynamic information to select routes.
- C. The changes in routing decisions are never reflected in the topology as well as traffic of the network.
- D. The main optimizing parameters related to this algorithm are hop count, distance and estimated transit time.

7. Which of the following is not a category of the adaptive routing algorithm?

- A. Centralized algorithm
- B. Isolation algorithm
- C. Distributed algorithm
- D. Flooding algorithm

8. The problem with the Flooding Non-Adaptive Routing Algorithm can be overcome with the help of_____

- A. Hop Count
- B. Sequence Number
- C. Spanning Tree
- D. All of these

9. Which of the following is not a drawback of the Broadcast routing technique?

- A. Inefficiency
- B. multicast
- C. unrealistic assumptions

- D. uncontrolled flooding
10. Following are the types of routing on MANET, except_____
- Proactive Routing
 - Reactive Routing
 - Hybrid Routing
 - Hyper Active Routing
11. Which of the following is not a routing metric?
- Signal stability
 - Choice of most followed route
 - Association stability
 - Route Length
12. Which of the following is true regarding Flat Routing Protocol?
- It is implemented by routers in which all routers are each other's peers.
 - It determines a route to a destination only when request for transmission to that node arrives.
 - Geographical information about a node is collected by another node using GPS mechanism.
 - The communication overhead is reduced but the delay is increased.
13. Which of the following is true regarding the wireless routing protocols?
- It was designed by Murthy and Garcia-Luna.
 - Here the route is chosen by selecting a neighbor node that would minimize the path cost.
 - It uses the properties of the distributed Bellman-Ford algorithm.
 - All the given options
14. In the Shortest Path Routing Algorithm, the cost of the link from one node to the other maybe a function of?
- Distance
 - Bandwidth
 - Average traffic between different nodes
 - All the given choices
15. Which of the following is not true regarding Dijkstra's algorithm?
- It works on the Directed Weighted Graph where each edge can have only positive weights.
 - It can have weights as in case of Bellman Ford Algorithm
 - It is a shortest path routing algorithm
 - All the given choices

Answer for Self Assessment

- | | | | | |
|-------|-------|-------|-------|-------|
| 1. A | 2. A | 3. C | 4. D | 5. D |
| 6. C | 7. D | 8. D | 9. D | 10. D |
| 11. B | 12. A | 13. D | 14. D | 15. B |

Review Questions

1. Discuss the role of network layer in the OSI model.
2. What are the main issues of concerns for the design of network layer?
3. Describe briefly how hierarchal algorithm works.
4. What is the main purpose of using router in a network?
5. Differentiate between:
 - a) Connectionless and connection-oriented service
 - b) Interior and Exterior Routing
 - c) Link state and distance vector routing

Further Reading



Andrew S. Tanenbaum, *Computer Networks*, Prentice Hall.

Behrouz A. Forouzan and Sophia Chung Fegan, *Data Communications and Networking*, McGraw-Hill Companies.

J. D. Spragins, *Telecommunications Protocols and Design*, Addison-Wesley.

Douglas Comer, *Computer Networks and Internets with Internet Applications*, 4th Edition, Prentice Hall.



<https://www.geeksforgeeks.org/multiple-access-protocols-in-computer-network>

Unit 10:Transport layer – Protocols

CONTANTS

Objectives

Introduction

10.1 Transport Layer

10.2 Services Provided by Transport Layer

10.3 Connection Oriented and Connectionless Services

10.4 Transport Layer Protocols

Summary

Self Assessment

Answer for self Assessment

Further Readings

Objectives

After this lecture, you would be able to

- understand the various services provided by the Transport Layer
- learn the various protocols used at the transport layer
- learn about the connection-oriented and connection-less services
- understand the nature of connection-oriented and connectionless services at both the layers
- understand the packet segment format of the UDP and the TCP protocol
- analyze the differences between the UDP and the TCP protocol

Introduction

The internet protocol is a part of the TCP IP protocol suite. The internet protocol is essential in assigning a unique IP address to a machine connected to the internet. Initially the IP addresses came in the form of IPv4 which was a 32 bit logical address which was represented as four octets of decimal numbers separated by colons these decimal numbers could take values between 0 to 255. However with extensive usage the maximum limit of the IP addresses is about to be reached. So to fulfill the need of a broader range for identifying numerous communication devices connected to the internet IPv6 was devised. It is a 128 bit hexadecimal address which can support up to 340 undecillion addresses. Further to support more number of computers in a network with limited IP addresses the concept of subnetting can overcome this limitation.

10.1 Transport Layer

The transport layer is a fourth layer on from the top of the OSI model. The main role of the transport layer is to provide the communication services directly to the application processes running on different hosts the transport layer provides a logical communication between application processes running on different hosts. Although the application

processes on different hosts are not physically connected the application processes use a logical communication provided by the transport layer to send the messages to each other, the transport layer protocols are implemented in the n systems but not in the network routers, a computer network provides more than one protocols to the network applications, For example, TCP and UDP or to transport layer protocols that provide a different set of services to the network layer, all transport layer protocols provide multiplexing or D multiplexing services. It also provides other services such as reliable data transfer bandwidth guarantee and delay guarantee, each of the applications in the application layer, have the ability to send a message by using TCP or UDP. Now the applications communicate by using either of these two protocols. So, both TCP and UDP will then communicate with the internet protocols in the internet layer. The applications can read and write to the transport layer, therefore we can see that communication is a two way process.

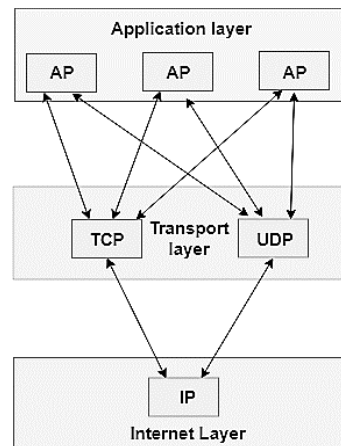


Figure 10. 1 TCP and UDP Communication in Transport Layer

The Figure 10.1 shown above, typically shows the various services being provided by the transport layer, and how the various communications are taking place between the application layer, and the Internet layer.

10.2 Services Provided by Transport Layer

The services provided by the transport layer are similar to those of the data link layer, the data link layer provides the services within a single network, while the transport layer provides the services across an entire network made up of many networks, the data link layer controls the physical layer while the transport layer controls all the lower layers, the services provided by the transport layer protocol can be divided into five categories. These are end-to-end delivery, addressing reliable delivery, flow control and multiplexing, you can very well see the different services provided by the transport layer with the help of this diagram. Now let us discuss each of the components of this diagram, one by one.

- 1) **End-to-end delivery**
- 2) **Reliable delivery**
- 3) **Flow control**
- 4) **Multiplexing**
- 5) **Addressing**

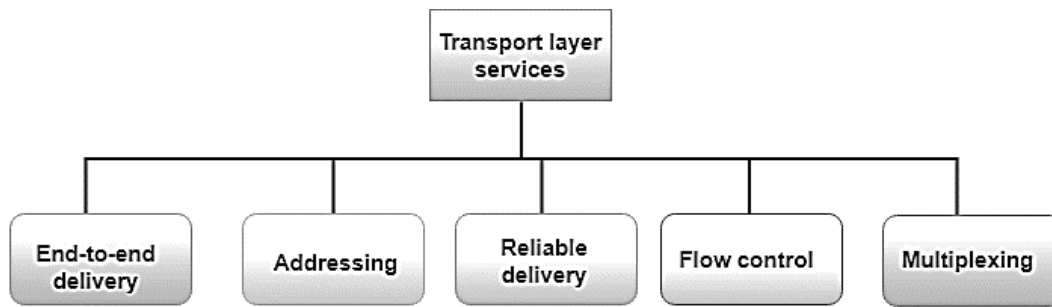


Figure 10. 2 Services Provided by Transport Layer

1. End-to-end delivery:

In end to end delivery, the transport layer transmits the entire message to the destination. Therefore, it ensures that the end to end delivery of an entire message from a source to the destination is perfectly done.

2. Reliable Delivery:

The second salient service provided by the transport layer is the reliable delivery, the transport layer provides reliable services by transmitting the lost and damaged packets, the reliable delivery has four aspects error control sequence control loss control and duplication control. Now let us see the various factors of reliable delivery, one by one

- a) Error control
- b) Sequence control
- c) Loss control
- d) Duplication control

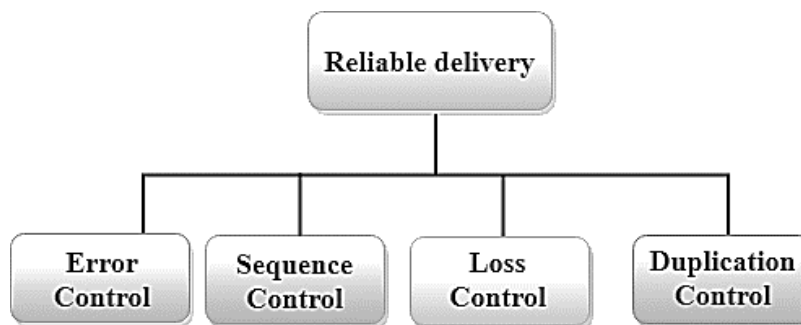


Figure 10. 3 Reliable Delivery Services Provided by Transport Layer

a) Error Control

Error control is the primary role of the transport layer. In reality, no transmission will be 100% error free while delivery. Therefore transport layer protocols are designed to provide a error free transmission, environment, the data link layer also provides the error handling mechanism but it ensures only node to node error free delivery. however, Node to Node reliability does not ensure the end to end reliability, the data link layer checks for the errors between each network. If an error is introduced inside one of the routers, then the error will not be caught by the data link layer, It only detects those errors that have been introduced between the beginning and the end of the link, therefore transport layer performs the checking from the errors, end to end, to ensure that the packets have arrived correctly at the destination. The Figure 10.3 is showing you how errors are not checked at the datalink layer, whereas the errors being checked at the data link layer, so how the different packets are being checked by the transport layer, and how the transport layer is performing the checking of these errors, to ensure that packets are arriving correctly at the destination.

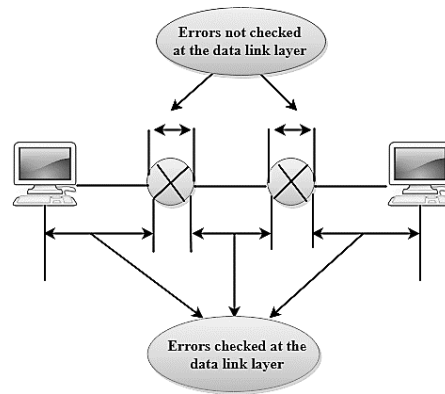


Figure 10. 4 Error Control in Transport Layer

b) Sequence Control

The second aspect of the reliability is the sequence control which is implemented at the transport layer on the sending end the transport layer is responsible for ensuring that the packets received from the upper layers can be used by the lower layers on the receiving end and ensures that the various segments of a transmission can be correctly reassembled.

c) Loss Control

Loss control is another aspect of reliability. Well the transport layer will ensure that all fragments of the transmission arrive at the destination and not some of them on the sending end all the fragments of transmission are given sequence numbers by a transport layer. These sequence numbers allow the receivers to transport layer to identify the missing segments.

d) Duplication Control

The next aspect that we will be talking about is duplication control. The transport layer guarantees that no duplicate data arrives at the destination sequence numbers are used to identify the lost packets. Similarly, it allows the receivers to identify and discard duplicate segments

e) Flow Control

Flow control is used to prevent the sender from overwhelming the receiver, if the receiver is overrun with too much data, then the receiver discards the packets and ask for retransmission of packets. The increased network congestion, and thus reducing the system performance. Now let us discuss about the flow control. This increases the network congestion and thus reduces the system performance. The transport layer is responsible for flow control. It uses the sliding window protocol that makes the data transmission more efficient, as well as it controls the flow of data so that the receiver does not become overwhelmed. The sliding window protocol is byte oriented rather than the frame oriented.

f) Multiplexing

The transport layer uses the multiplexing to improve the transmission efficiency multiplexing can occur in two ways. **Upward Multiplexing**

a) Downward Multiplexing

Well, what do they mean?let us see one by one.

a) Upward multiplexing:

by upward multiplexing we mean multiple transport layer connections use this same network connection to make more cost effective. So the transport layer sends several transmissions bound to the same destination along the same path. This is achieved through upward multiplexing. The diagram which is being depicted on the screen is showing you a typical example of the upward multiplexing. In the diagram you can see

how we are enabling a cost effective transport layer that is sending several transmissions bound to the same destination, following the same path.

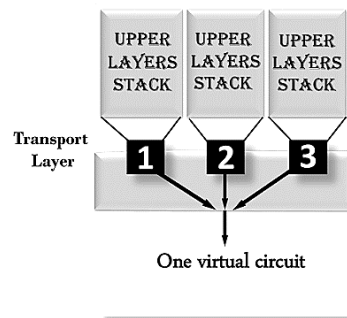


Figure 10. 5 Upward Multiplexing

b) Downward Multiplexing:

Downward multiplexing means one transport layer connection uses the multiple network connections. That means downward multiplexing allows the transport layer to split a connection among several paths to improve the throughput. This type of multiplexing is used when the network has a low or slow capacity. The Figure 10.6 which is being depicted on the screen, typically shows a typical example of a downward multiplexing. As you can very well see in the diagram, the transport layer splits a connection among several paths to improve the throughput.

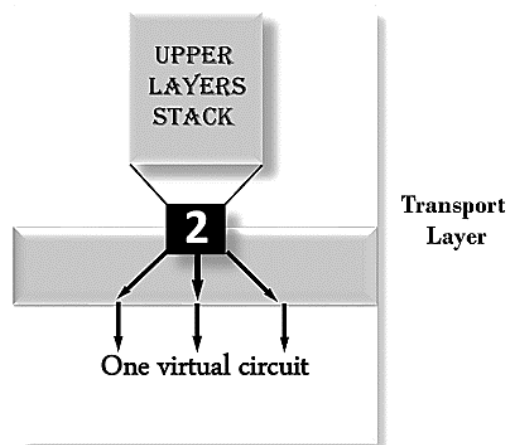


Figure 10. 6 Downward Multiplexing

3. Addressing

The next aspect that we shall be discussing is addressing. According to the layered model the transport layer interacts with the functions of the session layer. Many protocols combine sessions, presentation and application level protocols into a single layer which is known as the application layer. So, in these cases, delivery to the session layer means the delivery to the application layer, data generated by an application on one machine must be transmitted to the correct application on another machine, the transport layer provides a user address, which is specified as a station or a port. The port variable represents a particular TS user of a specific station where TS refers to the transport service, and the TSAP refers to the transport service access point each station has only one transport entity, the transport protocols needs to know which upper layer protocols are communicating. Now the Figure 10.7 is typically showing the various addressing mechanisms inside the transport layer, you can see that there are different upper level stacks which are able to communicate with each other, with the help of peer to peer communication on the transport layer.

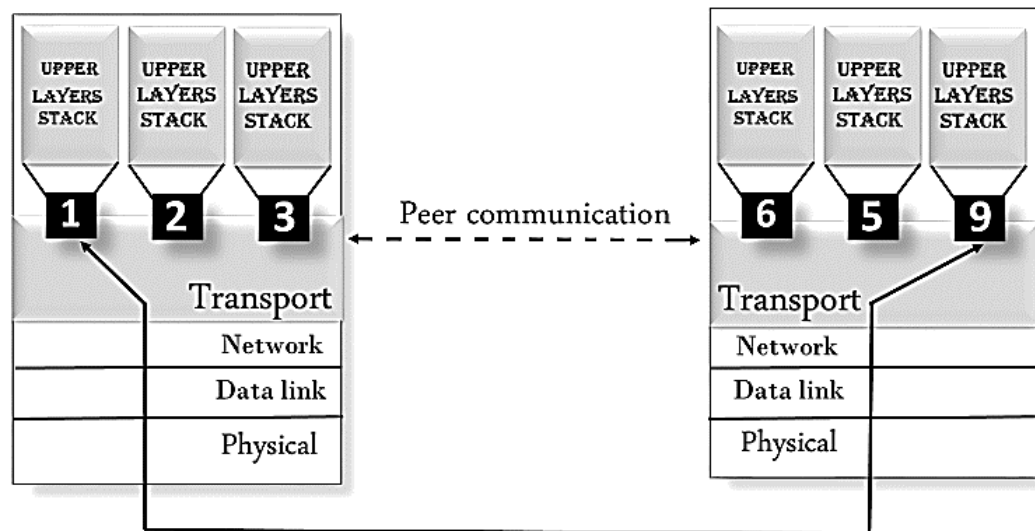


Figure 10. 7 Various addressing mechanisms inside the transport layer

10.3 Connection Oriented and Connectionless Services

Connection oriented and connection and services are the two data transmission services provided by the network layer protocols and the transport layer protocols. The connection-oriented services establish a connection, prior to sending the packets. That means, belonging to the same message from the source to the destination. On the other hand, the connectionless services consider each packet belonging to the same message as a different and independent entity and route them with a different path connection oriented and connectionless services show different behavior at the network layer, and the transport layer

1. What is Connection Oriented Service?

The connection-oriented service is related to the telephone system. It includes the connection establishment and connection termination. In connection-oriented service handshake method is used to establish the connection between the sender and the receiver.

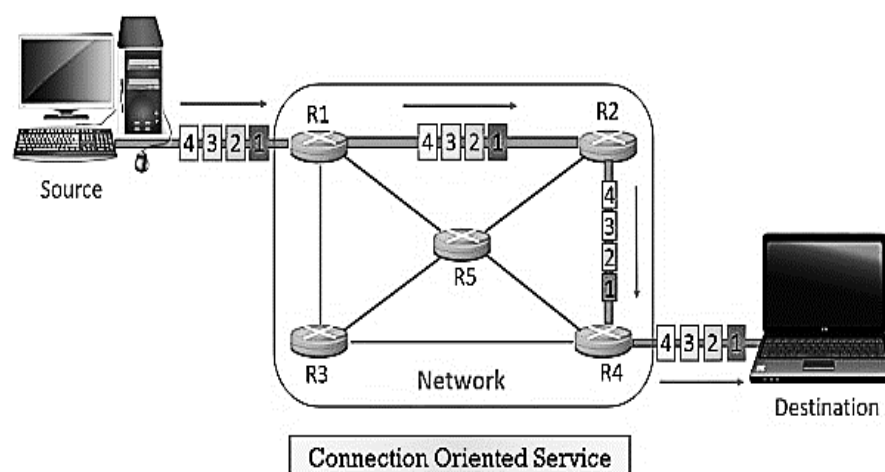


Figure 10. 8 Connection Oriented Service

The Figure 10.8 shows you a typical connection-oriented service, you can see the device one is being connected to device number two. And there are multiple paths to send our packets from device one to device two, but the selected path, which is being depicted with the blue dotted lines is

used to stream the data to understand this in a better elaborate way, let us consider the second diagram. As you can see in this diagram the data is being divided into sub packets that is packet number 123, and four, these packets are to be sent from the source to the destination. Now, you can see in the diagram, there are different routers which we can use. So, the routing decisions can be taken at different routers that is router number 1234, and five. So, according to the different parts, and the different parameters that we choose which could be the time, the cost, the accuracy, the turnaround time, we could decide upon which path we want to take up. So, in the current example we are picking up the path r one r two r four from the source to the destination. Also, you should note that the packets are getting transmitted in the same sequence that is packet number 1234, where a big message has been broken down into four packets which are sequenced and numbered. So, at the destination, the packets are again being received in the same sequence. Now the connection-oriented service first establishes a virtual connection between the source and the destination to establish a connection, a source sends a request packet to the destination. In response, the destination sends the acknowledgment packet to the source confirming that the destination is ready to accept the data from the source. It then transfers all data packets through the same dedicated the established connection. And after all packets are transferred it releases the connection. Meanwhile the router involved in the exchange of requests and acknowledgment packets between source and destination, define the virtual path that will be followed by all packets belonging to the same message. So we can see that the resources involved in data transfer are reserved before transferring all packets in a message, as all the data packets in the message follow the same path, their order is preserved as they reach the destination. After sending all data packets, the source sends a special baggage to terminate the connection. In response to it, the destination sends an acknowledgment confirming the termination of the connection, and all the routers, delete the path entry from the routing table.

Benefits Now let us see some benefits of the connection-oriented service as connection-oriented service provide acknowledgement at each action. It provides reliability in the service; there are fewer chances of packet loss as they travel up predefined path. Now the connection-oriented services are preferred over a long and steady conversation. As the virtual path is predefined, there are rare, or maybe no chances of congestion. In case of delay in data transmission. There is no delay in the transmission of packets, as there is a dedicated path for. Now, the TCP protocol is a connection-oriented protocol. This service works the same at both the network layer, and the transport layer that is first. It establishes a connection, then it exchanges data, and finally it terminates the connection, but its behavior slightly differs on both the layers. How does it differ. Let us see at the network layer, the connection oriented service is concerned regarding the coordination of source destination and routers involved in between the source and the destination, as there is coordination between source and destination, and all the routers in between and all the packets belonging to the same message, follow a dedicated established connection. So, we can implement flow control error control and congestion control in connection-oriented services. Now what happens at the transport layer let us see at the transport layer the connection-oriented service is concerned, only about the source and the destination. Here the packet show dependency on each other as all packets will go around the same allocated route. Now it is time for us to discuss what is connectionless service. Well, it is a method of data transmission between two computers in different networks. It is also termed to as a data gramme service. The service looks alike the postal system where each letter carries its source and destination address and each one of them is routed through a different path. The source is dividing the message into small acceptable packets and these packets are known as data grammes. So, the data grammes are eventually pushed into the network. Now, each data gramme may travel, the path that it suits, that means it can travel, a different path. Now, the network considers each data gramme or data packet as an independent entity that is no relationship is considered between these packets belonging to the same message. Each data gramme carries its source and destination address, and the router uses the destination address to route the data gramme to the destination, the packet received at the destination may be received out of order. Hence, the data grammes are assembled to recreate the original message.

2. What is Connectionless Service?

It is a method of data transmission between two computers in a different network. It is also termed as datagram service. This service look-alike the postal system where each letter carries its source & destination address and each one of them is routed through a different path. The

source divides the message into small acceptable packets these packets known as a datagram. These datagrams are individually pushed into the network. Each datagram may travel a different path. The network considers each datagram or data packet as an independent entity i.e. no relationship is considered between the packets belonging to the same message. Each datagram carries its source and destination address. The router uses the destination address to route the datagram to its destination. The packets received at the destination may be received out of order. Hence, the datagrams are assembled to recreate the original message.

Now to understand the connectionless service, let us have a look at Figure 10.9. As you can see, we are trying to send a message from the source to the destination. You can see that the message is being divided into four packets packet number 123, and four. Now, there are routers available through which we can make different routing decisions, the routers available are router number 1234, and five. Now, when you are sending the packets, there is no predefined path in the connectionless service. So, the packet can take any part which is most optimal at that point of time. But what is the drawback of this approach the packets are arriving, out of order at the destination. As you can see in the diagram, the packets are arriving, as packet number four to one and three.

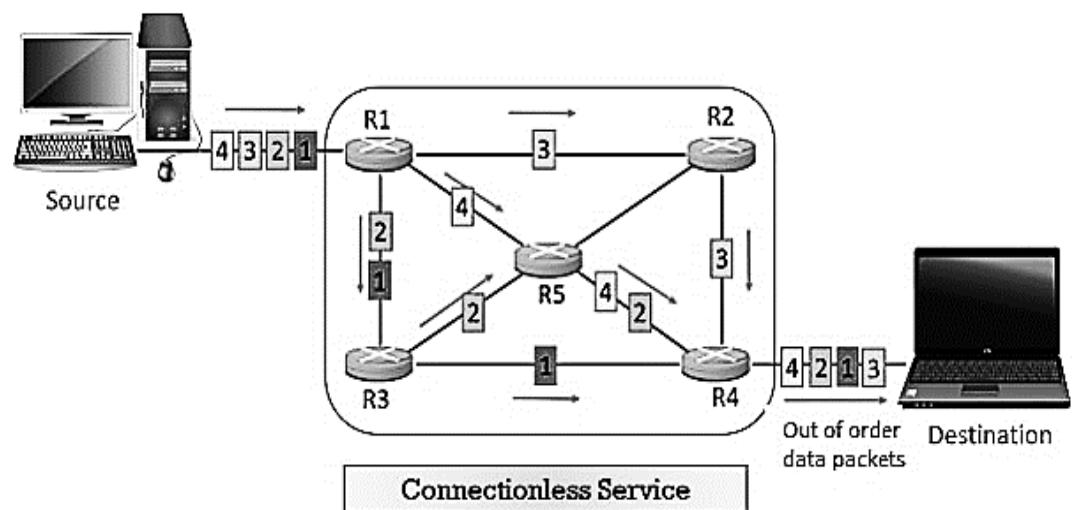


Figure 10. 9 Connectionless Service

The only overhead is now to rearrange these packets at the destination to receive the original message. So, it is related to the postal system, and it does not include any connection establishment and connection termination steps connectionless service does not give the guarantee, or reliability since the path is not predefined, and we can choose the best optimal path at the, at that particular point of time depending upon the different criteria's and conditions in this service, the packets do not follow the same path to reach the destination. So, the connectionless service is provided by the protocols of both network layer, as well as the transport layer. Though its basic function is the same that is it routes, each packet individually and independently over the network, but it may be through different data paths. However, it still behaves slightly different at both the network layer, and the transport layer. So, let us see how they behave differently

At Network Layer:

At the network layer, the connectionless service signifies different paths for different data packets belonging to the same message.

At Transport Layer:

at the transport layer the connectionless service exhibits independence between the packets, rather than the different parts that different packets belonging to the same message might follow, or will

follow. As the data packets belonging to the same message follow different paths, it may happen that they are received at the destination, out of order. Now, it can also be the case that one of the packet is lost. So what will happen at the transport layer, each packet is considered as an

independent entity and packets show no relationship with each other. So the destination transport layer will not even know that a packet has been lost. Here, we can conclude that we cannot implement flow control error control or congestion control in connection less service. So, the source divides the message into small acceptable packets, these packets are also known as data grammes. Now these data grammes are individually pushed into the network, and each data gramme may travel, a different part, the network considers each data gramme or data packet, as an individual entity that is there is no relation is considered between the packets belonging to the same message. Each data gramme carries its source and destination address. Now the router uses the destination address to route the data grammes to its destination.

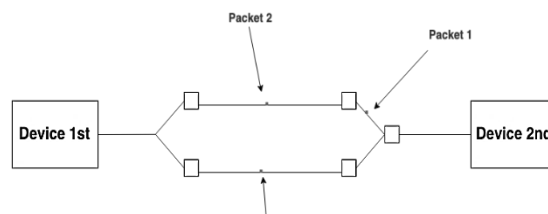


Figure 10. 10 Connectionless Service

3. Difference b/w Connection-Oriented & Connectionless Services

- Both Connection-oriented service and Connection-less service are used for the connection establishment between two or more than two devices. These type of services are offered by network layer.

Table 10. 1 Difference b/w Connection-Oriented & Connectionless Services

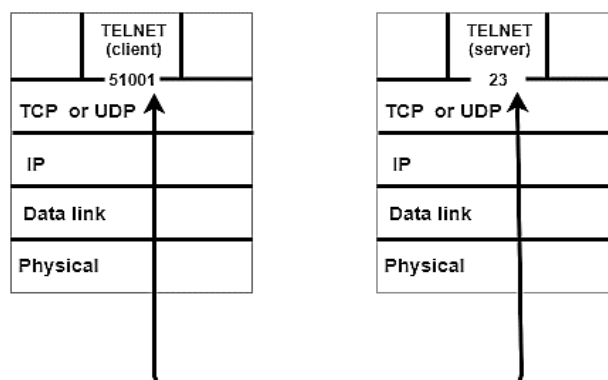
S.NO	Connection-oriented Service	Connection-less Service
1	Connection-oriented service is related to the telephone system.	Connection-less service is related to the postal system.
2	Connection-oriented service is preferred by long and steady communication.	Connection-less Service is preferred by bursty communication.
3	Connection-oriented Service is necessary.	Connection-less Service is not compulsory.
4	Connection-oriented Service is feasible.	Connection-less Service is not feasible.
5	In connection-oriented Service, Congestion is not possible.	In connection-less Service, Congestion is possible.
6	Connection-oriented Service gives the guarantee of reliability.	Connection-less Service does not give the guarantee of reliability.
7	In connection-oriented Service, Packets follow the same route.	In connection-less Service, Packets do not follow the same route.
8	Connection-oriented Services requires a bandwidth of high range.	Connection-less Service requires a bandwidth of low range.

4. Key Points to Remember

- Connection oriented service is based on the telephone system.
- Connection oriented service first establishes a connection between the source and destination.
- All packets belonging to the same message follows the same dedicated established connection to reach the destination in connection oriented service.
- At the network layer connection oriented service focuses on coordination between source, destination and all the routers between source and destination.
- At the transport layer connection, oriented service is only interested in source and destination. The service here is end to end.
- In connection oriented service we can implement flow control, error control & congestion control.
- Connectionless service is based on the postal services.
- Connectionless service considers each packet of the same message as a different and independent entity. Each data packet carries its source and destination address.
- In connectionless service, each packet of the same message may follow a different route to get delivered to the destination.
- In connectionless service, packets are routed based on the destination address on the packet.
- At the network layer, connectionless service signifies a different route for a different packet belonging to the same message.
- At the transport layer, connectionless service signifies the independency between the packets of the same message.
- In connectionless service, we cannot implement flow control, error control, and congestion control.

10.4 Transport Layer Protocols

The transport layer is represented by two protocols: TCP and UDP. The IP protocol in the network layer delivers a datagram from a source host to the destination host. Nowadays, the operating system supports multiuser and multiprocessing environments, an executing program is called a process. When a host sends a message to other host means that source process is sending a process to a destination process. The transport layer protocols define some connections to individual ports known as protocol ports. An IP protocol is a host-to-host protocol used to deliver a packet from source host to the destination host. Also, the transport layer protocols are port-to-port protocols that work on the top of the IP protocols to deliver the packet from the originating port to the IP services, and from IP services to the destination port. Each port is defined by a positive integer address, and it is of 16 bits.



1. UDP – User Datagram Protocol

UDP stands for **User Datagram Protocol**. It is a simple protocol and it provides non-sequenced transport functionality. It is a connectionless protocol. This type of protocol is used when reliability and security are less important than speed and size. It is an end-to-end transport level protocol that adds transport-level addresses, checksum error control, and length information to the data from the upper layer. The packet produced by the UDP protocol is known as a user datagram.

User Datagram Format

The user datagram has a 16-byte header which is shown below:

Source port address 16 bits	Destination port address 16 bits
Total Length 16 bits	Checksum 16 bits
Data	

Where,

- **Source port address:** It defines the address of the application process that has delivered a message. The source port address is of 16 bits address.
- **Destination port address:** It defines the address of the application process that will receive the message. The destination port address is of a 16-bit address.
- **Total length:** It defines the total length of the user datagram in bytes. It is a 16-bit field.
- **Checksum:** The checksum is a 16-bit field which is used in error detection.

Disadvantages of UDP Protocol

- UDP provides basic functions needed for the end-to-end delivery of a transmission.
- It does not provide any sequencing or reordering functions and does not specify the damaged packet when reporting an error.
- UDP can discover that an error has occurred, but it does not specify which packet has been lost as it does not contain an ID or sequencing number of a particular data segment.

2. TCP - Transmission Control Protocol

TCP stands for Transmission Control Protocol. It provides full transport layer services to applications. It is a connection-oriented protocol means the connection established between both the ends of the transmission. For creating the connection, TCP generates a virtual circuit between sender and receiver for the duration of a transmission.

Features of TCP Protocol

1. Stream data transfer:

- TCP protocol transfers the data in the form of contiguous stream of bytes.
- TCP group the bytes in the form of TCP segments and then passed it to the IP layer for transmission to the destination.
- TCP itself segments the data and forward to the IP.

2. Reliability:

- TCP assigns a sequence number to each byte transmitted and expects a positive acknowledgement from the receiving TCP.

- If ACK is not received within a timeout interval, then the data is retransmitted to the destination.
 - The receiving TCP uses the sequence number to reassemble the segments if they arrive out of order or to eliminate the duplicate segments.
3. **Flow Control:**
- When receiving, TCP sends an acknowledgement back to the sender indicating the number the bytes it can receive without overflowing its internal buffer.
 - The number of bytes is sent in ACK in the form of the highest sequence number that it can receive without any problem. This mechanism is also referred to as a window mechanism.
4. **Multiplexing:**
- Multiplexing is a process of accepting the data from different applications and forwarding to the different applications on different computers.
 - At the receiving end, the data is forwarded to the correct application. This process is known as demultiplexing.
 - TCP transmits the packet to the correct application by using the logical channels known as ports.
5. **Logical Connections:**
- The combination of sockets, sequence numbers, and window sizes, is called a logical connection.
 - Each connection is identified by the pair of sockets used by sending and receiving processes.
6. **Full Duplex:**
- TCP provides Full Duplex service, i.e., the data flow in both the directions at the same time.
 - To achieve Full Duplex service, each TCP should have sending and receiving buffers so that the segments can flow in both the directions.
 - TCP is a connection-oriented protocol. Suppose the process A wants to send and receive the data from process B. The following steps occur:
Establish a connection between two TCPs.
Data is exchanged in both the directions.
The Connection is terminated.

3. TCP Segment Format

Source port address 16 bits				Destination port address 16 bits				
Sequence number 32 bits								
Acknowledgement number 32 bits								
HLEN 4 bits	Reserved 6 bits	U R G	A C K	P S H	R S T	S Y N	F I N	Window size 16 bits
Checksum 16 bits				Urgent pointer 16 bits				
Options & padding								

Where,

- **Source port address:** is used to define the address of the application program in a source computer. It is a 16-bit field.
- **Destination port address:** is used to define the address of the application program in a destination computer. It is a 16-bit field.
- **Sequence number:** A stream of data is divided into two or more TCP segments.
- The 32-bit sequence number field represents the position of the data in an original data stream.
- **Acknowledgement number:** A 32-bit acknowledgement number acknowledges the data from other communicating devices. If ACK field is set to 1, then it specifies the sequence number that the receiver is expecting to receive.
- **Header Length (HLEN):** It specifies the size of the TCP header in 32-bit words. The minimum size of the header is 5 words, and the maximum size of the header is 15 words. Therefore, the maximum size of the TCP header is 60 bytes, and the minimum size of the TCP header is 20 bytes.
- **Reserved:** is a six-bit field which is reserved for future use.
- **Control bits:** Each bit of a control field functions individually and independently. A control bit defines the use of a segment or serves as a validity check for other fields.

There are total six types of flags in control field:

- i) **URG:** urgent.
 - ii) **ACK:** acknowledgement number.
 - iii) **PSH:** data must be pushed with higher throughput.
 - iv) **RST:** reset the TCP connection
 - v) **SYN:** Synchronize the sequence numbers
 - vi) **FIN:** The sender has finished sending data.
-
- i) **URG:** The URG field indicates that the data in a segment is urgent.
 - ii) **ACK:** When ACK field is set, then it validates the acknowledgement number.
 - iii) **PSH:** The PSH field is used to inform the sender that higher throughput is needed so if possible, data must be pushed with higher throughput.
 - iv) **RST:** The reset bit is used to reset the TCP connection when there is any confusion occurs in the sequence numbers.
 - v) **SYN:** The SYN field is used to synchronize the sequence numbers in three types of segments: connection request, connection confirmation (with the ACK bit set), and confirmation acknowledgement.
 - vi) **FIN:** The FIN field is used to inform the receiving TCP module that the sender has finished sending data. It is used in connection termination in three types of segments: termination request, termination confirmation, and acknowledgement of termination confirmation.

The various fields shared in FIN are:

- a) Window Size
 - b) Checksum
 - c) Urgent pointer
 - d) Options and padding
-
- a) **Window Size:** The window is a 16-bit field that defines the size of the window.
 - b) **Checksum:** The checksum is a 16-bit field used in error detection.

- c) **Urgent Pointer:** If URG flag is set to 1, then this 16-bit field is an offset from the sequence number indicating that it is a last urgent data byte.
- d) **Options and Padding:** It defines the optional fields that convey the additional information to the receiver.

4. Differences between TCP & UDP

Table 10. 2Differences Between TCP & UDP

Basis of Comparison	TCP	UDP
Definition	TCP establishes a virtual circuit before transmitting the data.	UDP transmits the data directly to the destination computer without verifying whether the receiver is ready to receive or not.
Connection Type	It is a Connection-Oriented protocol	It is a Connectionless protocol
Speed	slow	High
Reliability	It is a reliable protocol.	It is an unreliable protocol.
Header size	20 bytes	8 bytes
Acknowledgement	It waits for the acknowledgement of data and has the ability to resend the lost packets.	It neither takes the acknowledgement, nor it retransmits the damaged frame.

Summary

The main role of the transport layer is to provide the communication services directly to the application processes running on different hosts the transport layer provides a logical communication between application processes running on different hosts. Although the application processes on different hosts are not physically connected the application processes use a logical communication provided by the transport layer to send the messages to each other, the transport layer protocols are implemented in the n systems but not in the network routers, a computer network provides more than one protocols to the network applications, For example, TCP and UDP or to transport layer protocols that provide a different set of services to the network layer, all transport layer protocols provide multiplexing or the multiplexing services.

Keywords

Flow Control: It is used to prevent the sender from overwhelming the receiver, if the receiver is overrun with too much data, then the receiver discards the packets and ask for retransmission of packets.

Error Control: It is the primary role of the transport layer. In reality, no transmission will be 100% error free while delivery. Therefore, transport layer protocols are designed to provide a error free transmission, environment, the data link layer also provides the error handling mechanism but it ensures only node to node error free delivery.

Upward Multiplexing: The multiple transport layer connections use this same network connection to make more cost effective. So, the transport layer sends several transmissions bound to the same destination along the same path. This is achieved through upward multiplexing.

Downward Multiplexing: By downward multiplexing, we mean that one transport layer connection uses the multiple network connections. That means downward multiplexing allows the transport layer to split a connection among several paths to improve the throughput. This type of multiplexing is used when the network has a low or slow capacity.

Connection Oriented Service: The connection-oriented services establish a connection, prior to sending the packets.

Connectionless Services: The connectionless services consider each packet belonging to the same message as a different, and independent entity and route them with a different path

UDP: UDP or the User Datagram Protocol is a simple connectionless protocol that provides non sequence transport functionality.

TCP. The Transmission Control Protocol (TCP) provides full transport layer services to the applications. It is a connection oriented protocol which means that the connection establishment between both the ends of the transmission is being handled by this protocol for creating the connection that TCP generates a virtual circuit between a sender and the receiver for the duration of a transmission.

Self Assessment

Select the correct answer for the following questions

- The main aspects related to reliable delivery in the transport layer are.
 - Error control
 - Sequence control
 - Loss control
 - Duplication control
 - A and C
 - A and D
 - B and C.
 - A, B, C and D.
- Which of the following is true regarding upward multiplexing?
 - It means multiple transport layer connections use different network connections.
 - For cost-effectiveness, the transport layer sends several transmissions bound for the same destination along the same path
 - It allows the transport layer to split a connection among several paths to improve the throughput.
 - This type of multiplexing is used when networks have a low or slow capacity.
- Which of the following is not true regarding the Addressing function of the Transport Layer?
 - The transport layer provides the user address which is specified as a station or port.
 - Each station has only one transport entity.
 - The transport layer protocols need not know which upper-layer protocols are communicating.
 - The port variable represents a particular TS user of a specified station known as a Transport Service access point (TSAP).
- Which of the following is not true regarding the Addressing function of the Transport Layer?
 - The transport layer provides the user address which is specified as a station or port.
 - Each station has only one transport entity.
 - The transport layer protocols need not know which upper-layer protocols are communicating.
 - The port variable represents a particular TS user of a specified station known as a Transport Service access point (TSAP).

5. Which of the following is not true regarding flow control feature of the Transport Layer?
 - A. The underlying rule of flow control is to maintain a synergy between a fast process and a slow process.
 - B. It uses the sliding window protocol that makes the data transmission more efficient
 - C. The transport layer controls the flow of data so that the receiver does not become overwhelmed.
 - D. Sliding window protocol is frame oriented rather than byte oriented.
6. Which of the following is true regarding connection-oriented service?
 - A. Connection-oriented service is related to the telephone system.
 - B. It includes the connection establishment and connection termination.
 - C. Handshake method is used to establish the connection between sender and receiver.
 - D. All the given choices
7. Which of the following is not a benefit of the connection-oriented service?
 - A. It provides reliability in the service.
 - B. There are fewer chances of packet loss
 - C. It is preferred by bursty communication.
 - D. It provides a long and steady conversation.
8. Which of the following is not true regarding connectionless service?
 - A. In connection-less service, packets follow the same route.
 - B. Connection-less service requires a bandwidth of low range.
 - C. Connection-less service is related to the postal system.
 - D. In connection-less service, congestion is possible.
9. Which of the following is not true?
 - A. In connection-oriented service we can implement flow control, error control & congestion control.
 - B. In connectionless service, each packet of the same message should follow the same route to get delivered to the destination.
 - C. Connectionless service considers each packet of the same message as a different and independent entity.
 - D. In connectionless service, packets are routed based on the destination address on the packet.
10. Which of the following are transport layer protocols used in networking?
 - A. TCP and FTP
 - B. UDP and HTTP
 - C. TCP and UDP
 - D. HTTP and FTP
11. Transmission control protocol?
 - A. Is a connection-oriented protocol
 - B. Uses a three-way handshake to establish a connection
 - C. Receives data from application as a single stream
 - D. All the given choices
12. Which of the following is false with respect to UDP?
 - A. Connection-oriented
 - B. Unreliable
 - C. Transport layer protocol

- D. Low overhead
13. In context of the TCP segment format, the various fields shared in FIN are
- Window Size
 - Checksum
 - Options and padding
 - All the given choices
14. With respect to the TCP Segment Format, what is the Window size?
- It is a 8-bit field which defines the size of the window
 - It is a 16-bit field which defines the size of the window
 - It is a 12-bit field which defines the size of the window
 - It is a 32-bit field which defines the size of the window
15. Which of the following is true regarding the TCP Segment format?
- It transmits the data directly to the destination computer without verifying whether the receiver is ready to receive or not.
 - It is a connectionless protocol
 - It establishes a virtual circuit before transmitting the data.
 - It has a header size of 8 bytes

Review Questions

- How is transport layer different from data link layer when the services provided at both the layers are almost similar?
- Why transport layer is required when both the network and transport layers provide connectionless and connection oriented services?
- Why UDP is used when it provides unreliable connectionless service to the transport layer?
- What is the purpose of flow control?
- Describe the TCP and its major advantages over UDP.

Answer for self Assessment

- | | | | | |
|------|------|------|------|------|
| 1 D | 2 B | 3 C | 4 C | 5 D |
| 6 D | 7 C | 8 A | 9 B | 10 C |
| 11 D | 12 A | 13 D | 14 B | 15 C |



Further Readings

- Achyut S Godbole and Atul Kahate published, *Web Technologies*, Tata McGraw Hill.
- Andrew S. Tanenbaum, *Computer Networks*, Prentice Hall.
- Behrouz A. Forouzan and Sophia Chung Fegan, *Data Communications and Networking*, McGraw-Hill Companies.
- Douglas Comer, *Computer Networks and Internets with Internet Applications*, 4th Edition, Prentice Hall.
- J. D. Spragins, *Telecommunications Protocols and Design*, Addison-Wesley.



<https://www.geeksforgeeks.org/multiple-access-protocols-in-computer-network>

Unit 11: Introduction to Computer Networks

CONTENTS

Objectives

Introduction

11.1 What is Congestion?

11.2 Causes/Costs of Congestion

11.3 Traffic Profiles

11.4 Congestion Control

11.5 Quality of Service (QoS)

11.6 Need for QoS

11.7 Integrated Services (IntServ)

Summary

Keywords

Self-Assessment

Review Questions

Further Readings

Objectives

After this lecture, you would be able to:

- learn the basic concept of congestion control
- understand the various causes of congestion
- learn the various ways of congestion control
- learn the impact of congestion on Quality of Service
- understand the different flow characteristics and the need of QoS
- learn the various types of QoS requirements and QoS solutions

Introduction

Congestion is any stage, into which we aren't able to carry on the normal communication. As you can see on the screen. This is a typical example of traffic congestion, where there is converging traffic. And at any particular point of time there is a choking situation in the traffic. So the traffic is not having a smooth flow and this is what we mean by congestion. So, it is the reduced quality of service that occurs when a network node or link is carrying more data than it can handle. So when will a traffic overflow occur. A traffic overflow would only occur when a particular road is carrying traffic more than its own capacity. If it is carrying traffic, more than the capacity. It is only under such circumstances that an overflow or a congestion situation can occur. Now typical effects of this congestion include queuing delay. That means you will have long waiting in the queue packet loss packets might be lost during transmission or blocking of new connections, so new connections will be blocked, they will not be able to traverse on that particular network. So, this is about congestion. So, what is a consequence of congestion, well, a consequence of congestion is that an incremental increase in offered load, leads, either only to a small increase, or even a decrease in the network throughput. So, the performance of the network depreciates. Now, let us discuss the basic requirements, or the quality of service specifications. Now, quality of service requirements can be classified or specified as belonging to the four factors. So, quality of service can be specified as delay. Delay variation which I've already told you it is called jitter. It could be on the basis of throughput, that that is the output that I get after a particular time period, and the error rate which is there, and there are two types of quality of service solutions, which I, which are available for us.

One is the stateless solution. Another is the stateful solution. So what are these, let us see one by one. In stateless solution routers maintain no fine-grained state about traffic. So one positive factor of it is that it is scalable, and it is robust, but it has weak services, as there is no guarantee about the kind of delay or performance in a particular application, which we have to encounter, talking about the integrated services, or the in itself, what exactly do we mean by that. Let us now see architecture for providing quality of service guarantees in IP networks for individual application sessions. So the architecture, which is to provide quality of service guarantee, and in, in the IP network because it is the Internet Protocol networks that we use so individual application sessions need to take care of this. It relies on the resource reservation and routers need to maintain state information of allocated resources.

11.1 What is Congestion?

Congestion is the reduced quality of service that occurs when a network node or link is carrying more data than it can handle. It is a state occurring in network layer. When the message traffic is so high that it slows down the network response time. We can understand the concept of congestion as too many sources sending too much data fast for network to handle. So, it is a crucial problem and we need to handle this situation. Typical effects include queuing delay, packet loss or the blocking of new connections. A consequence of congestion is that an incremental increase in offered load, leads either only to a small increase or even a decrease in network throughput.

1. Effects of Congestion

There are different effects of congestion, like the buffer overflow at the routers might happen. It's just like a mother feeding a child. The mother waits for the child to nod his head, and then she gives the next spoon of soup or, or the meal that they are having. Now when the child gives the nod, he agrees to take the next bite/sip. Similar is the case of the networks, but in case of networks. If there is no such mechanism as congestion control, there are chances that there might be overflow buffers and data might get lost. So, there is another problem which is called long delays. And this happens because of going in the router buffers. So as the delay increases the performance is going to decrease. So, delay is something which is not desired. It has reverse repercussions on the performance of the system. And if the delay increases, the retransmission occurs, making situation even worse, because if we are not able to handle the first packet and then because of timeout. We have to retransmit, another copy of the packet, thus increasing the traffic load on the network, and even more. Lost packets. (Buffer overflow at routers)

11.2 Causes/Costs of Congestion:

The causes and costs of the

1. Scenario 1

This scenario says that there are two senders and there are two receivers. And they all follow one single route. So, we require an infinite buffer that means we can handle as much data as we get, and there is no retransmission. So, this is the ideal situation. There is no retransmission or infinite buffer which is very unrealistic. We cannot have infinite buffers, and we need to retain the packets which are being sent. So, what are the various aspects that are required to be taken care of? Figure shows two things, which are throughput and delay. We say that throughput increases with load. Throughput is basically the time it takes for you to respond or to complete that task. Now as the load increases throughput and the maximum total load will also increase. Since there are two senders, so the maximum load would be C divided by two, because there were two sessions is now large delays, when congested lead to the load which is stochastic, so the load in this case is stochastic load, and we need to deal with it.

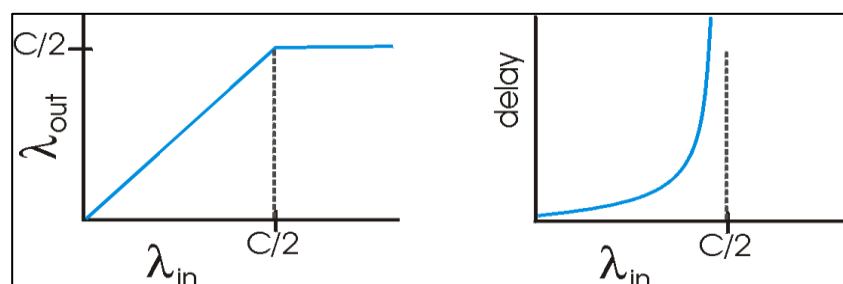


Figure 11.1 Scenario 1 of Congestion Control

2. Scenario 2

In the second scenario, there is one route, which is having only finite buffers. As in the very first example, there were infinite buffers but, in this case, there are finite buffers which can send or retransmit the packets. The Lost packets that means there are chances that packet might be lost. Now it sounds to be realistic. In the previous example, it was not realistic that infinite buffer is there, and no retransmission is required that means, that is the ideal situation in which no packet is lost, and a kid can retain as much spoons as is given to him. It is just like the kid having infinite capacity in his mouth, and he can retain each and everything that anybody is giving him to eat. But that is not the real-world scenario. So, there should be a finite buffer. If a packet is lost, it needs to retransmit the packet which was not there in Scenario number one. Now, its time to discuss a few things. It should be seen that the throughput is coming in (throughput in), and the throughput which is going out (throughput out) should be equal. If it is always equal, we say that the throughput is good. So, to maximize the good output, it should be ensured that the system should be able to handle all the traffic that comes in. So, the incoming and outgoing traffic is handled without any breaks and the throughput is the ideal throughput (In-throughput is equal to Out-throughput). This is the ideal and the best condition. Now let's understand what perfect retransmission is? It means to retransmit only when a loss condition is there. When the input is large enough, and the output is not that large, that means somebody has fed the kid 10 spoons, but he has only taken the five spoons. So, where have the five spoons gone. Maybe you have spilled it off on the floor. So those are lost.

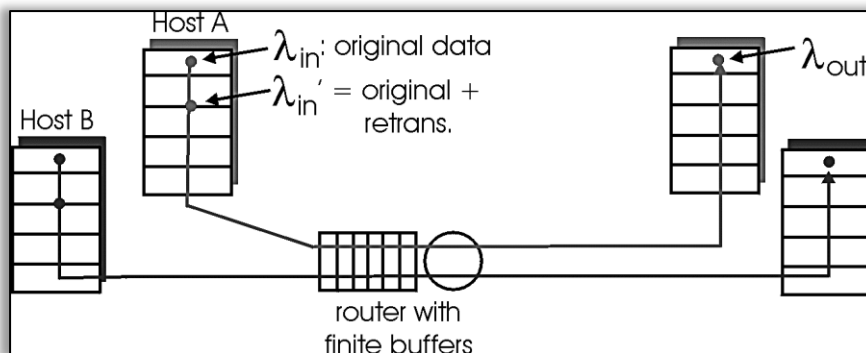


Figure 11 2Senario2 of Congestion Control

Always: (goodput)

- $\lambda_{in} = \lambda_{out}$ to maximize goodput!

“perfect” retransmission:

- retransmit only when loss: $\lambda'_{in} > \lambda_{out}$

The last scenario that is basically not be perfect transmission, there is a need to retransmit and the mother(sender) again need to give the five spoons to the kid(receiver) so that it could be handled. Now actual retransmission of delays involves no loss packet. That means if the sender is retransmitting the packets and there is a delay, then there is no loss, and the packets get delivered. Also, it makes larger than the perfect case for the same throughput. So, whatever is the throughput, there is a larger and the perfect case for the same throughput. So, under the ideal condition we should be having equal or more output compared to the inputs. But it's not possible to have more output than the input, because feasibly, it's not possible that sender has not sent those many packets than those received by the receiver. But yes, there is an ideal situation in which throughput-in is equal to the throughput-out. This is called good put.

- Actual retransmission of delayed (not lost) packet
- Makes λ'_{in} larger (than perfect case) for same λ_{out}

Q: what happens as λ_{in} and λ'_{in} increase ?

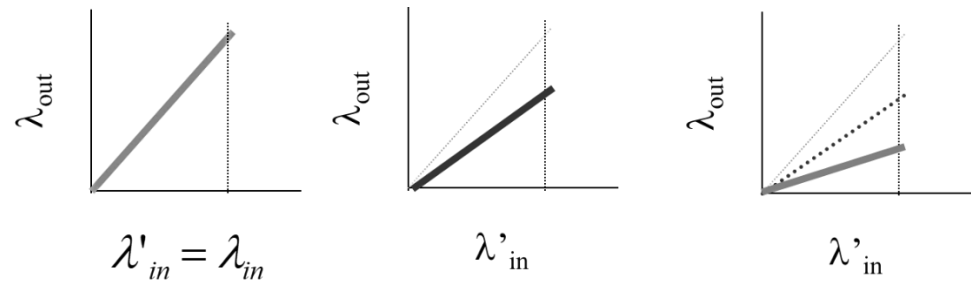
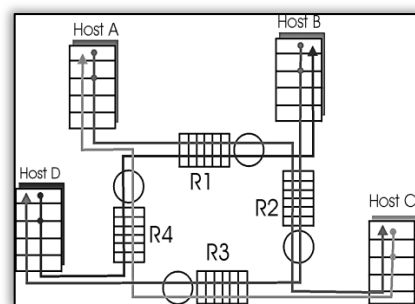
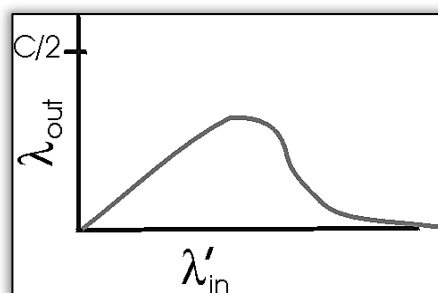
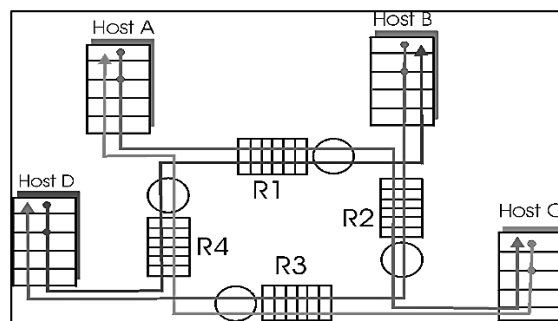


Fig 11.4.: “costs” of congestion

The figure shows that the variations in the costs of congestion and an attempt is being made to find out how this cost varies with different factors. So, the retransmission leads to multiple copies of the packets being transmitted (more overhead) thus reducing the throughput. Retransmission means repeatedly doing the same task because it was not properly done at a particular point of time. The throughput is reduced considerably, because in that particular time period, the system will not be able to produce those many outputs due to retransmissions.

3. Scenario 3

To understand the third scenario, let us assume that there are four centers, and there is a multi-hop path. In the previous scenarios, there was only one path between the host and the destination, and this was not the ideal situation. So, buffers will be required, and there was a chance that the packets could be lost. Now complicating the situation even more, let us assume that there are four senders. So, host A, host B, host C, and hosts D are the four senders. We have multiple multi hop path, which means there are multiple paths for sending the data. So we can go via router number one to four. There are different routers and timeout and retransmission might occur. So what happens as the throughput, or the speed of the data, which is coming inside increases? Does it lead to increase in the output? Well the Answer to this is NO. It doesn't happen. For example if a mother is in a hurry and she has to leave for shopping and wants to feed the kid. So, she gives the bytes to the kid. Since she is in a hurry, so she tries to increase the speed to two times, then to three times, four times and so on. So, as the speed is going to increase. It's not that your output will also increase at a particular point.



4. Another “cost” of congestion:

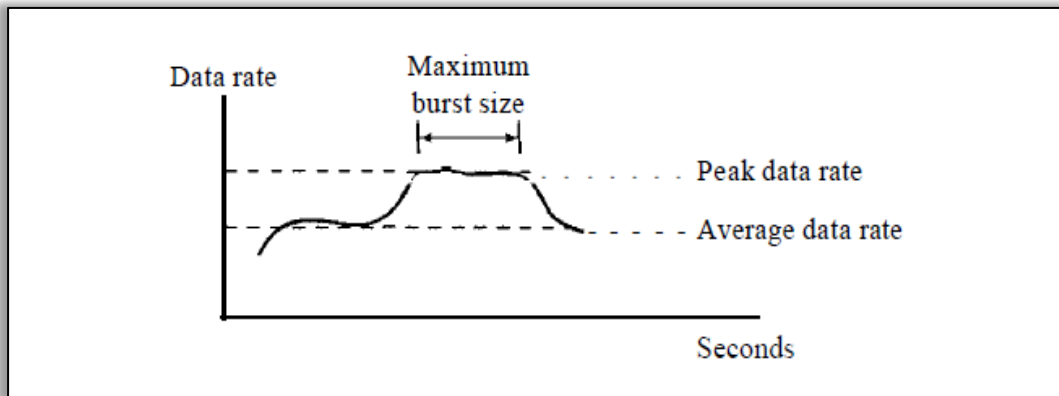
when packet dropped, any “upstream” transmission capacity used for that packet was wasted!

Data Traffic

- The main focus of congestion control and quality of service is data traffic.
- In congestion control we try to avoid traffic congestion.
- In quality of service, we try to create an appropriate environment for the traffic.

Traffic Descriptor

Traffic descriptors are qualitative values that represent a data flow.



Where

5. Average Data Rate: The average data rate is the number of bits sent during a period of time, divided by the number of seconds in that period. We use the following equation:

6. Peak Data Rate: The peak data rate defines the maximum data rate of the traffic. In the figure, it is the maximum y axis value. It is a very important measurement because it indicates the peak bandwidth that the network needs for traffic to pass through without changing its data flow.

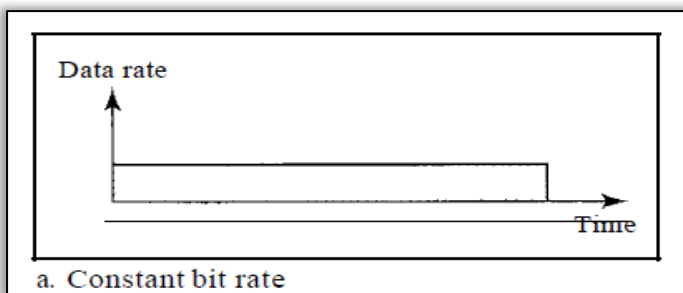
7. Maximum Burst Size: Although the peak data rate is a critical value for the network, it can usually be ignored if the duration of the peak value is very short. The maximum burst size normally refers to the maximum length of time the traffic is generated at the peak rate.

Effective Bandwidth: The effective bandwidth is the bandwidth that the network needs to allocate for the flow of traffic. The effective bandwidth is a function of three values: average data rate, peak data rate, and maximum burst size. The calculation of this value is very complex.

11.3 Traffic Profiles

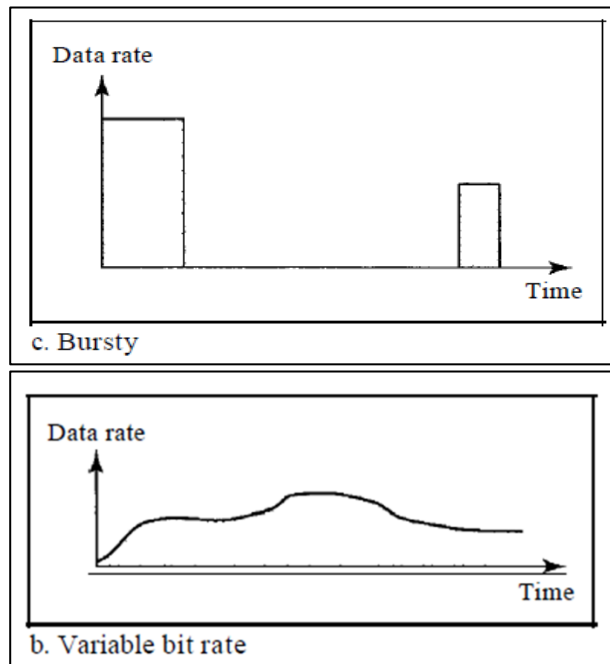
1. Constant Bit Rate:

A constant-bit-rate (CBR), or a fixed-rate, traffic model has a data rate that does not change. In this type of flow, the average data rate and the peak data rate are the same. The maximum burst size is not applicable. This type of traffic is very easy for a network to handle since it is predictable. The network knows in advance how much bandwidth to allocate for this type of flow.



2. Variable Bit Rate:

In the variable-bit-rate (VBR) category, the rate of the data flow changes in time, with the changes smooth instead of sudden and sharp. In this type of flow, the average data rate and the peak data rate are different. The maximum burst size is usually a small value.



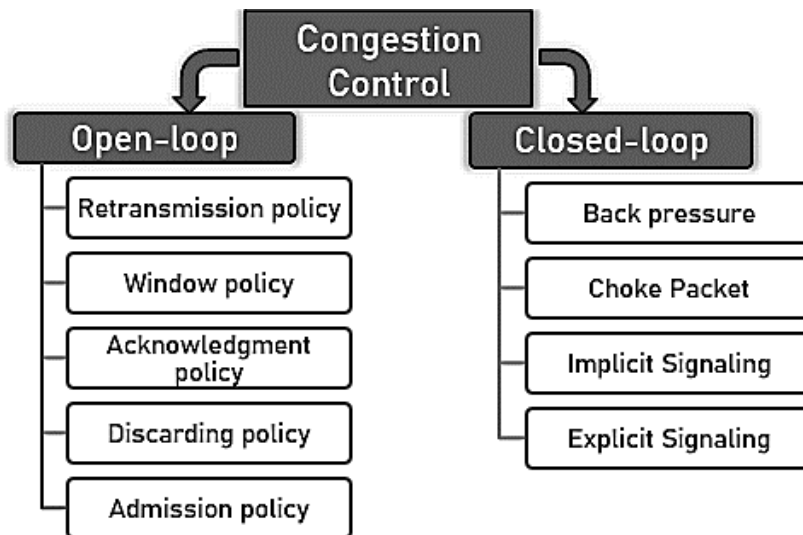
3. Bursty Data Rate:

In this category the data rate changes suddenly in a very short time. It may jump from zero, for example, to 1 Mbps in a few microseconds and vice versa. It may also remain at this value for a while. The average bit rate and the peak bit rate are very different values in this type of flow.

11.4 Congestion Control

An important issue in a packet-switched network is congestion. Congestion in a network may occur if the load on the network-the number of packets sent to the network-is greater than the capacity of the network-the number of packets a network can handle. Congestion control refers to the mechanisms and techniques to control the congestion and keep the load below the capacity. Congestion control refers to techniques and mechanisms that can either prevent congestion, before it happens, or remove congestion, after it has happened. In general, we can divide congestion control mechanisms into two broad categories:

- a) Open-loop congestion control (prevention)
- b) Closed-loop congestion control (removal).



1. Open Loop Congestion Control

Retransmission Policy

Here if the sender feels that a sent packet is lost or corrupted, the packet needs to be retransmitted. The retransmission policy and the retransmission timers must be designed to optimize efficiency and at the same time prevent congestion.

Window Policy

The type of window at the sender may also affect congestion. The Selective Repeat window is better than the Go-Back-N window for congestion control. In the Go-Back-N window, when the timer for a packet times out, several packets may be resent, although some may have arrived safe and sound at the receiver.

Acknowledgement Policy.

If the receiver does not acknowledge every packet it receives, it helps prevent congestion.

Discarding Policy:

In this policy less sensitive packets may be discarded when congestion is likely to happen

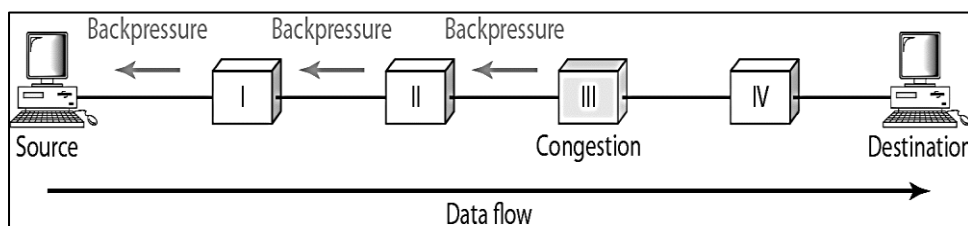
Admission Policy

A router can deny establishing a virtual circuit connection if there is congestion in the network or if there is a possibility of future congestion.

2. Closed Loop Congestion Control

Backpressure

It is a node-to-node congestion control that starts with a node and propagates, in the opposite direction of data flow, to the source. The backpressure technique can be applied only to virtual circuit networks, in which each node knows the upstream node from which a flow of data is coming.



Choke Packet:

A choke packet is a packet sent by a node to the source to inform it about congestion. The warning is from the router, which has encountered congestion, to the source station directly.

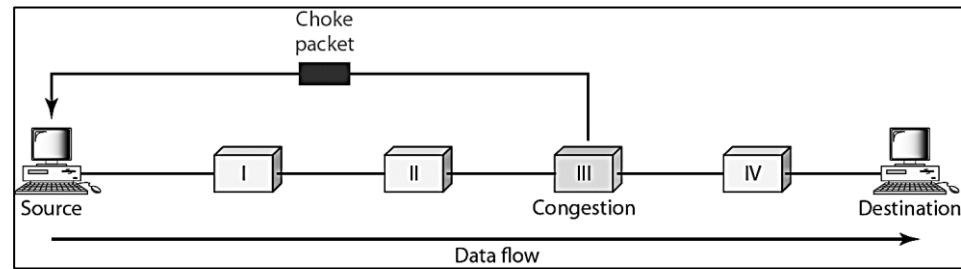


Figure: Choke Packet:

Implicit Signaling:

There is no communication between the congested node or nodes and the source. The source guesses that there is congestion somewhere in the network from other symptoms.

Explicit Signaling:

The node that experiences congestion can explicitly send a signal to the source or destination. It can further be of two types:

- a) Backward Signaling
- b) Forward Signaling

a) Backward Signaling

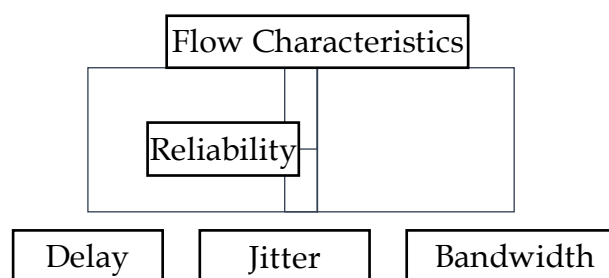
Backwards signaling means a bit can be sent in a packet moving in the direction opposite to the congestion. This bit can warn the source that there is congestion and that it needs to slow down to avoid the discarding of packets. So, that is why it is called backwards because it is in the opposite direction. So, this bit can warn the source that there is congestion and that it needs to slow down to avoid the discarding of packets. So, a backward signal is formed. A bit can be sent in the opposite direction to state that there is a congestion packets should not be send.

b) Forward Signaling

A bit can be set in a packet moving in the direction of the congestion. This bit can warn the destination that there is congestion goes in the direction of the flow. So a bit can be sent in a packet moving in the direction of the congestion that is where it is called forward. So this bit can warn the destination that there is congestion so, this way. There are different approaches by which we come to know that there are different congestion conditions.

11.5 Quality of Service (QoS)

It refers to traffic control mechanisms that seek to either differentiate performance based on application or network-operator requirements or provide predictable or guaranteed performance to applications, sessions or traffic aggregates. The basic phenomenon for QoS means in terms of packet delay and losses of various kinds. In other words Quality of Service can be defined as something a flow seeks to attain.



11.6 Need for QoS

The video and audio conferencing require bounded delay and loss rate. The video and audio streaming requires bounded packet loss rate, it may not be so sensitive to delay. Time-critical applications (real-time control) in which bounded delay is considered to be an important factor. Valuable applications should be provided better services than less valuable applications.

1. QoS Specification

QoS requirements can be specified as:

- Delay
- Delay Variation (Jitter)
- Throughput
- Error Rate

2. QoS Solutions

There are two types of QoS Solutions:

- a) Stateless Solutions
- b) Stateful Solutions

a) Stateless Solutions

In stateless solutions Routers maintain no fine grained state about traffic, one positive factor of it is that it is scalable and robust. But it has weak services as there is no guarantee about kind of delay or performance in a particular application which we have to encounter.

b) Stateful Solutions

In stateful solutions Routers maintain per flow state as flow is very important in providing the Quality-of-Service. i.e. providing powerful services such as guaranteed services and high resource utilization, provides protection and is much less scalable and robust.

11.7 Integrated Services (IntServ)

Architecture for providing QoS guarantees in IP networks for individual application session's. Relies on resource reservation, and routers need to maintain state information of allocated resources and respond to new call setup requests. Network decides whether to admit or deny a new call setup request.

1. IntServ QoS Components

- Resource reservation: call setup signaling, traffic, QoS declaration, per-element admission control.
- QoS-sensitive scheduling e.g. WFQ queue discipline.
- QoS-sensitive routing algorithm(QSPF)
- QoS-sensitive packet discard strategy.
- RSVP-Internet Signaling

It creates and maintains distributed reservation state, initiated by the receiver and scales for multicast, needs to be refreshed otherwise reservation times out as it is in soft state. Latest paths discovered through "PATH" messages (forward direction) and used by RESV messages (reserve direction).

2. Call Admission

Session must first declare it's QoS requirement and characterize the traffic it will send through the network. R-specification: defines the QoS being requested, i.e. what kind of bound we want on the delay, what kind of packet loss is acceptable, etc. T-specification: defines the traffic characteristics like bustiness in the traffic. A signaling protocol is needed to carry the R-spec and T-spec to the routers where reservation is required. Routers will

admit calls based on their R-spec, T-spec and based on the current resource allocated at the routers to other calls.

Differentiated Service

Differentiated Service is a stateful solution in which each flow doesn't mean a different state. It provides reduced state services i.e. maintain state only for larger granular flows rather than end-to-end flows tries to achieve best of both worlds. Intended to address the following difficulties with IntServ and RSVP:

1. Flexible Service Models
2. Simpler signaling
1. Flexible Service Models:

IntServ has only two classes that aim for the following,

- want to provide more qualitative service classes.
- want to provide 'relative' service distinction.
- Streaming Live Multimedia

2. Simpler Signaling:

Many applications and users may only want to specify a more qualitative notion of service.



Example:

- Internet radio talk show
- Live sporting event.

Streaming:

playback buffer, playback buffer can lag tens of seconds after and still have timing constraint.

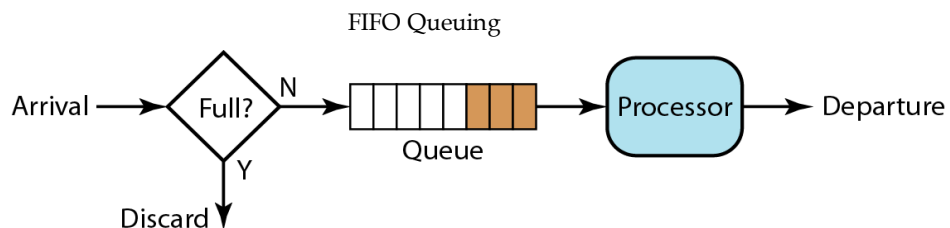
Interactivity:

Fast forward is impossible, but rewind and pause is possible.

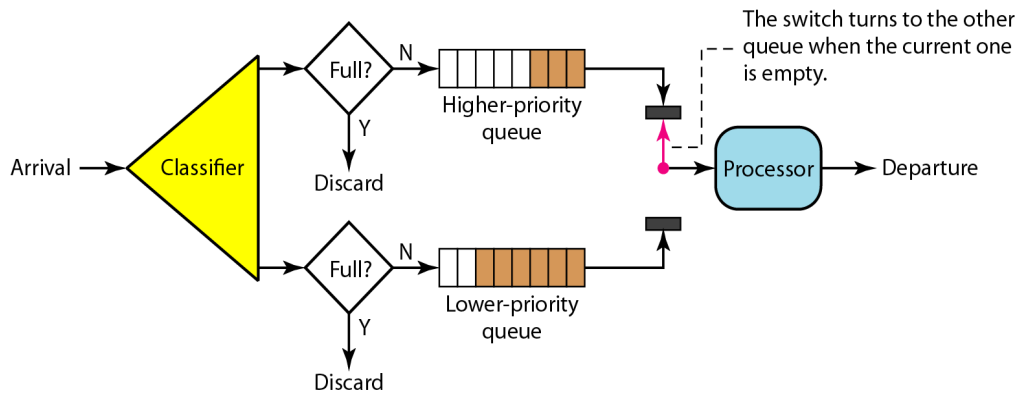
Techniques to Improve QoS

- a) Scheduling
- b) Traffic Shaping
- c) Resource Reservation
- d) Admission Control

a) Scheduling

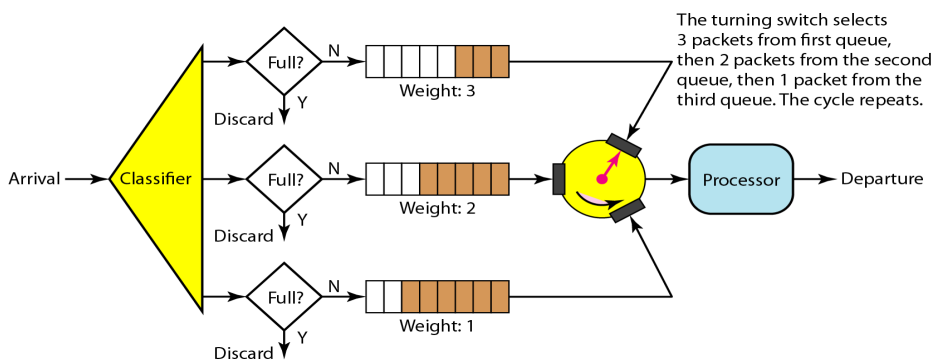


Priority Queuing



Weighted Fair Queuing

The turning switch selects 3 packets from first queue, then 2 packets from the second queue, then 1 packet from the third queue. The cycle repeats



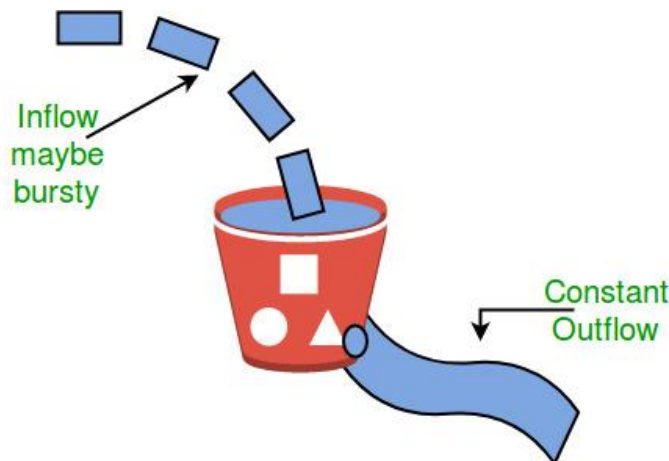
11.8 Congestion Control Algorithms

- Leaky Bucket Algorithm
- Token bucket Algorithm

Leaky Bucket Algorithm

Let us consider an example to understand

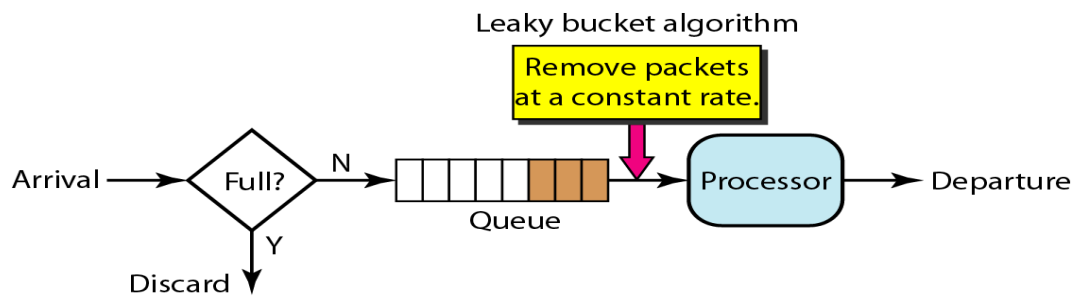
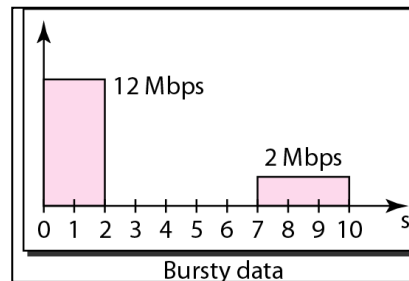
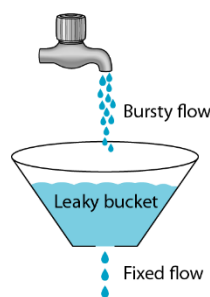
- Imagine a bucket with a small hole in the bottom.
- No matter at what rate water enters the bucket, the outflow is at constant rate.
- When the bucket is full with water additional water entering spills over the sides and is lost.



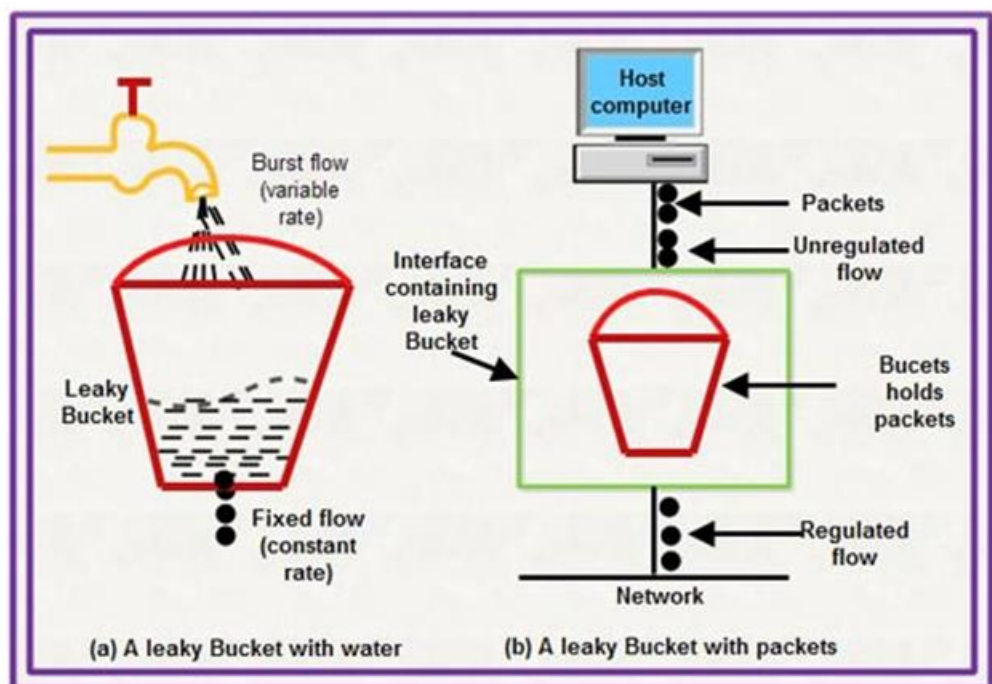
Similarly, each network interface contains a leaky bucket and the following steps are involved in leaky bucket algorithm:

- When host wants to send packet, packet is thrown into the bucket.
- The bucket leaks at a constant rate, meaning the network interface transmits packets at a constant rate.
- Bursty traffic is converted to a uniform traffic by the leaky bucket.
- In practice the bucket is a finite queue that outputs at a finite rate.

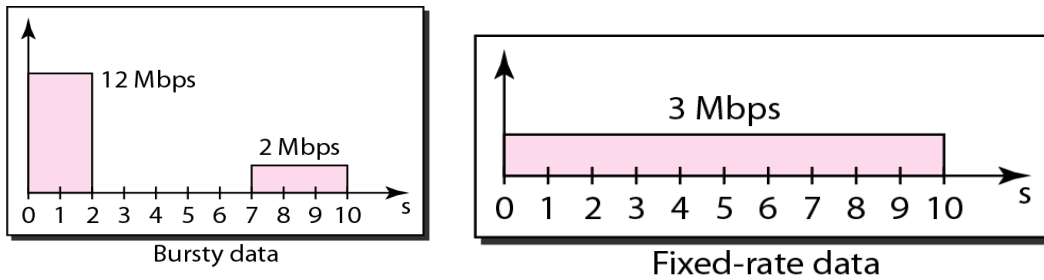
11.9 Traffic Shaping



Leaky Bucket



In the figure, we assume that the network has committed a bandwidth of 3 Mbps for a host.



The use of the leaky bucket shapes the input traffic to make it conform to this commitment.

In the figure the host sends a burst of data at a rate of 12 Mbps for 2s, for a total of 24 Mbits of data.

The host is silent for 5s and then sends data at a rate of 2Mbps for 3s, for a total of 6Mbits of data.

In all, the host has sent 30 Mbits of data in 10s.

The leaky bucket smooths the traffic by sending out data at a rate of 3Mbps during the same 10s.

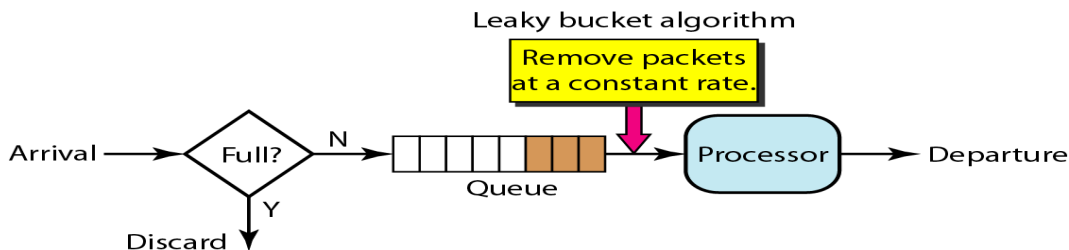
Without the leaky bucket, the beginning burst may have hurt the network by consuming more bandwidth than is set aside for this host.

We can also see that the leaky bucket may prevent congestion. As an analogy, consider the freeway during rush hour (busty traffic). If, instead, commuters could stagger their working hours, congestion on our freeways could be avoided.



A leaky bucket algorithm shapes burst traffic into fixed-rate traffic by averaging the data rate. It may drop the packets if the bucket is full.

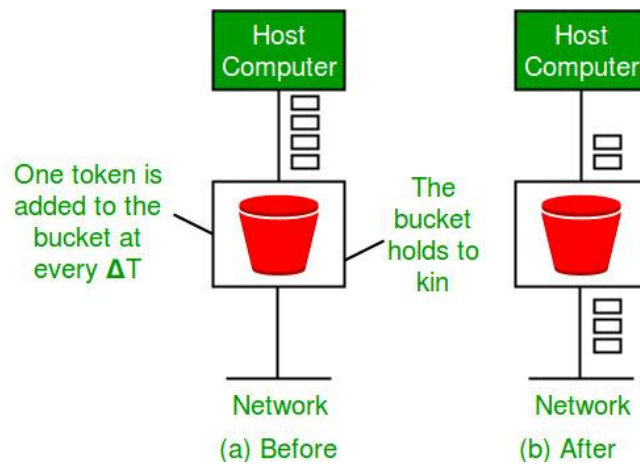
Implementation of Leaky Bucket Algorithm



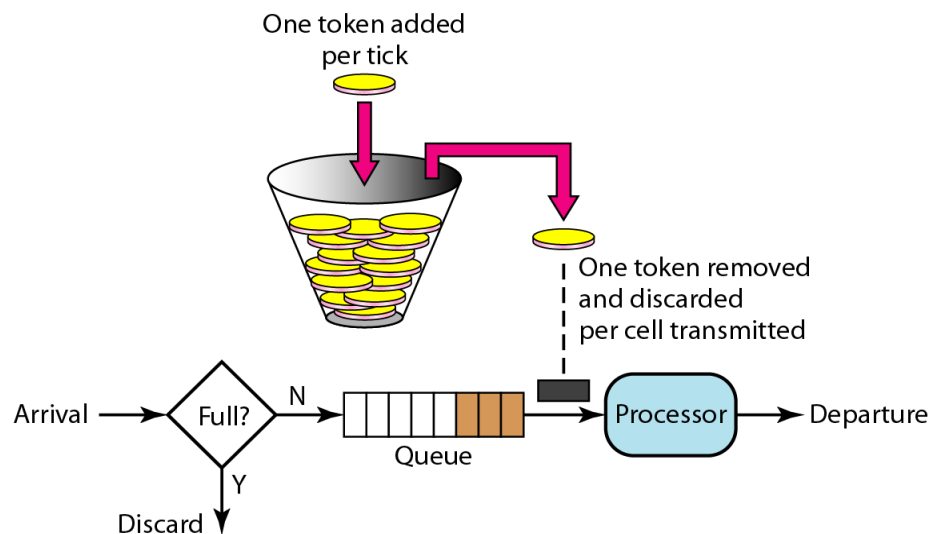
Token bucket Algorithm

Need of token bucket Algorithm:-

- The leaky bucket algorithm enforces output pattern at the average rate, no matter how bursty the traffic is.
- So in order to deal with the bursty traffic we need a flexible algorithm so that the data is not lost.
- One such algorithm is token bucket algorithm.



- The leaky bucket is very restrictive. It does not credit an idle host.
- For example, if a host is not sending for a while, its bucket becomes empty.
- Now if the host has busy data, the leaky bucket allows only an average rate.
- The time when the host was idle is not taken into account.



The token bucket algorithm allows idle hosts to accumulate credit for the future in the form of tokens.

For each tick of the clock, the system sends n tokens to the bucket.

The system removes one token for every cell (or byte) of data sent.

For example, if n is 100 and the host is idle for 100 ticks, the bucket collects 10,000 tokens. Now the host can consume all these tokens in one tick with 10,000 cells, or the host takes 1000 ticks with 10 cells per tick.

In other words, the host can send bursty data as long as the bucket is not empty. Figure shows the idea. The token bucket can easily be implemented with a counter.

The token is initialized to zero. Each time a token is added, the counter is incremented by 1. Each time a unit of data is sent, the counter is decremented by 1. When the counter is zero, the host cannot send data.

Steps of Token bucket Algorithm

Steps of this algorithm can be described as follows:

- In regular intervals tokens are thrown into the bucket. f
- The bucket has a maximum capacity. f

- If there is a ready packet, a token is removed from the bucket, and the packet is sent.
- If there is no token in the bucket, the packet cannot be sent.

Ways in which token bucket is superior to leaky bucket:

- The leaky bucket algorithm controls the rate at which the packets are introduced in the network, but it is very conservative in nature.
- Some flexibility is introduced in the token bucket algorithm. In the token bucket, algorithm tokens are generated at each tick (up to a certain limit).
- The leaky bucket algorithm controls the rate at which the packets are introduced in the network, but it is very conservative in nature.
- Some flexibility is introduced in the token bucket algorithm. In the token bucket, algorithm tokens are generated at each tick (up to a certain limit).
- For an incoming packet to be transmitted it must capture a token and the transmission takes place at the same rate.
- Hence some of the busy packets are transmitted at the same rate if tokens are available and thus introduces some amount of flexibility in the system.
- Ways in which token bucket is superior to leaky bucket:
- For an incoming packet to be transmitted it must capture a token and the transmission takes place at the same rate.
- Hence some of the busy packets are transmitted at the same rate if tokens are available and thus introduces some amount of flexibility in the system.

Token bucket Algorithm

Formula: $M * s = C + \rho * s$

where S - is time taken

M - Maximum output rate

ρ - Token arrival rate

C - Capacity of the token bucket in byte

Resource Reservation

A flow of data needs resources such as a buffer, bandwidth, CPU time, and so on. The quality of service is improved if these resources are reserved beforehand.

Admission Control

Admission control refers to the mechanism used by a router, or a switch, to accept or reject a flow based on predefined parameters called flow specifications.

Summary

Congestion is any stage, into which we aren't able to carry on the normal communication. As you can see on the screen. This is a typical example of traffic congestion, where there is converging traffic. And at any particular point of time there is a choking situation in the traffic. So the traffic is not having a smooth flow and this is what we mean by congestion. So, it is the reduced quality of service that occurs when a network node or link is carrying more data than it can handle. So when will a traffic overflow occur. A traffic overflow would only occur when a particular road is carrying traffic more than its own capacity. If it is carrying traffic, more than the capacity. It is only under such circumstances that an overflow or a congestion situation can occur. Now typical effects of this congestion include queuing delay. That means you will have long waiting in the queue packet loss packets might be lost during transmission or blocking of new connections, so new connections will be blocked, they will not be able to traverse on that particular network. So, this is about congestion. So, what is a consequence of congestion, well, a consequence of congestion, is that an incremental increase in offered load, leads, either only to a small increase, or even a decrease in the network throughput. So, the performance of the network depreciates. Now, quality of service requirements can be classified or specified as belonging to the four factors. So, quality of service can be specified as delay. Delay variation which I've already told you it is called jitter. It could be on the basis of

throughput, that that is the output that I get after a particular time period, and the error rate which is there, and there are two types of quality of service solutions, which I, which are available for us. One is the stateless solution. Another is the stateful solution. So what are these, let us see one by one. In stateless solution routers maintain no fine-grained state about traffic. So one positive factor of it is that it is scalable, and it is robust, but it has weak services, as there is no guarantee about the kind of delay or performance in a particular application, which we have to encounter, talking about the integrated services, or the in itself, what exactly do we mean by that. Let us now see architecture for providing quality of service guarantees in IP networks for individual application sessions. So, the architecture, which is to provide quality of service guarantee, and in, in the IP network because it is the Internet Protocol networks that we use so individual application sessions need to take care of this. It relies on the resource reservation and routers need to maintain state information of allocated resources.

Keywords

Stateless Solutions - In stateless solutions Routers maintain no fine grained state about traffic, one positive factor of it is that it is scalable and robust. But it has weak services as there is no guarantee about kind of delay or performance in a particular application which we have to encounter.

Stateful Solutions - In stateful solutions Routers maintain per flow state as flow is very important in providing the Quality-of-Service. i.e. providing powerful services such as guaranteed services and high resource utilization, provides protection and is much less scalable and robust..

A constant-bit-rate (CBR), or a fixed-rate, traffic model has a data rate that does not change. In this type of flow, the average data rate and the peak data rate are the same. Local Area Network: A LAN is a form of local (limited distance), shared packet network for computer communications.

Leaky Bucket Algorithm: It is a network with varioable input but output at a constant data rate

Self-Assessment

1. Congestion in a network or internetwork occurs because routers and switches have _____.
 - A. Tables
 - B. Queues
 - C. Cross points
 - D. None of the given choices
2. In a network, when the load is much less than the capacity of the network, the delay is _____.
 - A. at a maximum
 - B. at a minimum
 - C. constant
 - D. none of the given choices
3. In a network, when the load reaches the network capacity, the delay _____.
 - A. increases sharply
 - B. decreases sharply
 - C. remains constant
 - D. cannot be predicted
4. In a network, when the load is below the capacity of the network, the throughput _____.
 - A. increases sharply
 - B. increases proportionally with the load
 - C. declines sharply.
 - D. declines proportionately with the load.
5. In _____ congestion control, mechanisms are used to alleviate congestion after it happens

- A. Open-loop
 - B. Closed-loop
 - C. Both open-loop and closed-loop
 - D. neither open-loop nor closed-loop
6. The technique of _____ refers to a congestion control mechanism in which a congested node stops receiving data from the immediate upstream node or nodes
- A. backpressure
 - B. choke packet
 - C. implicit signaling
 - D. explicit signaling
7. A _____ is a packet sent by a node to the source to inform it of congestion.
- A. backpressure
 - B. implicit signaling
 - C. choke packet
 - D. explicit signaling
8. In _____, there is no communication between the congested node or nodes and the source. The source guesses that there is congestion somewhere in the network from other symptoms
- A. backpressure
 - B. implicit signaling
 - C. choke packet
 - D. explicit signaling
9. In the _____ method, the signal is included in the packets that carry data.
- A. backpressure
 - B. implicit signaling
 - C. choke packet
 - D. explicit signaling
10. Congestion is the reduced quality of service that occurs when a network node or link is carrying_____
- A. more data than it can handle
 - B. less data than it can handle
 - C. data to the server
 - D. All the given choices
11. Which of the following is an effect of congestion?
- A. queuing delay
 - B. packet loss
 - C. blocking of new connections
 - D. All the given choices
12. In _____, we try to create an appropriate environment for the traffic.
- A. congestion control
 - B. quality of service
 - C. both the given choices
 - D. None of these

13. Which of the following is not a requirement of Quality of Service?
- A. Jitter
 - B. Throughput
 - C. Error Rate
 - D. Stateless solution
14. Which of the following is not true regarding stateless solutions?
- A. In stateless solutions Routers maintain no fine-grained state about traffic
 - B. It is scalable
 - C. It is robust.
 - D. It has strong service
15. Which of the following is not true regarding Leaky Bucket Algorithm?
- A. It shapes burst traffic into fixed-rate traffic by averaging the data rate.
 - B. It enforces output pattern at the average rate, no matter how bursty the traffic is.
 - C. It never drops the packets even if the bucket is full.
 - D. It is not perfect when dealing with bursty traffic

State whether the following is true or false:

16. The quality of service (QoS) of computer networks is evaluated with respect to the traffic priority.
17. Bandwidth has no role to play in providing a good quality of service.
18. The best effort traffic model handles all Internet requests with equal priority and serves them with the first come first serve strategy.
19. The congestion management tool may include priority queuing, custom queuing, weighted air queuing, etc.
20. The link fragmentation and interleave process segment small packet into large packets interleaving the voice packet.
21. Shaping is used to prevent the overflow problem in buffers by limiting the full bandwidth potential of the packets of applications.

Answers: Self-Assessment

- | | | | |
|-----------|----------|----------|-----------|
| 1. B | 2. B | 3. A | 4. B |
| 5. B | 6. A | 7. C | 8. B |
| 9. D | 10. A | 11. D | 12. B |
| 13. D | 14. D | 15. C | 16. True |
| 17. False | 18. True | 19. True | 20. False |
| 21. True | | | |

Review Questions

1. Explain the general principles of congestion.
2. What do you understand by QoS? Describe the basic QoS structure.
3. Discuss the following two algorithms:
 - (a) Leaky Bucket
 - (b) Token Bucket
4. What are two types of congestion control? Where is congestion control implemented in each case?
5. Discuss the various causes of the costs of Congestion.

Further Reading



Andrew S. Tanenbaum, *Computer Networks*, Prentice Hall

Behrouz A. Forouzan and Sophia Chung Fegan, *Data Communications and Networking*, McGraw-Hill Companies

Burton, Bill, *Remote Access for Cisco Networks*, McGraw-Hill Osborne Media McGraw-Hill Osborne Media

Rajneesh Agrawal and Bharat Bhushan Tiwari, *Computer Networks and Internet*, Vikas Publication

William A Shay, *Understanding Communication and Networks*, 3rd Edition, Thomson Press.



<https://www.geeksforgeeks.org/basics-computer-networking/>

Unit 12: Application Layer – Services and Protocols

CONTENTS

Objectives

Introduction

12.1 Telnet (Terminal Network)

12.2 FTP

12.3 Domain Name System

12.4 Organization of Domain

12.5 Simple Mail Transfer Protocol

12.6 POP Protocol

12.7 Internet Message Access Protocol (IMAP)

Summary

Keywords

Self-Assessment

Review Questions

Answers: Self Assessment

Further Readings

Objectives

After this lecture, you would be able to:

- learn about the TERminal NETwork and the various tasks performed by it.
- understand about the working and the tasks performed by the File Transfer Protocol
- learn about the working, structure and benefit of Domain Name System
- learn concepts of DNS security and understand the challenges of DNS cache poisoning
- understand the significance, types and use of Simple Mail Transfer Protocol
- learn the different types of SMTP models and the basic SMTP commands
- learn about the Post Office Protocol and the Internet Message Access Protocol
- understand the way email gets transmitted in the POP protocol
- learn the various aspects related to the IMAP protocol

Introduction

The upper three layers namely session, presentation and application layers are considered as user or application layers of the OSI models. They are implemented in software. In most of the protocols, the functions of these layers are converged into a single layer called the application layer. TCP is one of the examples of such types of protocols. The application layer, the highest layer of OSI model interacts with software applications, which enable source and destination machines to communicate properly. It provides different services, which are described herein.

12.1 Telnet (Terminal Network)

Telnet is an application layer protocol, which can be used on the internet, or the local area network. It provides a bi-directional interactive text-oriented communication service by using virtual terminal connection. It is basically a client server protocol, which is based upon a reliable connection-oriented transport system. That means, that focus is on reliability in case of the Telnet protocol. Also, it is important to note that it uses port number 23. This port number is used to establish the connection with the transmission control protocol does Telnet is a client server application that allows a user to log on to a remote machine, and lets the user to access any application program on a remote computer. In case a manager is sitting back at your home and might be operating your computer which is in your office with the help of telnet. Now tell it uses the network virtual terminal system to encode the characters on the local system on the server side which is the remote machine, the network Virtual Terminal and decodes the characters to form an acceptable to the remote machine. So, it has to be converted into form which acceptable to the remote machine. To do so, the network Virtual Terminal is used. Telnet is a protocol that provides a bidirectional eight-bit byte-oriented communication facility. That means, the communication can take place in both directions and the data can be sent and received. There are many application protocols which are built upon the Telnet protocol. It is a client application which is intended for use with the Telnet protocol and is included by default with most operating systems. However, installation of a telnet client package may be needed to install it with the Microsoft Windows desktop operating systems and some Linux distributions.

When working on a Windows desktop system, if the Telnet Client cannot be found. What should be too. So, it can generally be added by going into "Turn Windows features on or off" in the Windows control panel and selecting the Telnet Client. On Windows Server, it is in the feature's summary selection of the Server Manager. So, a user needs to go to the Server Manager, and find out the feature summary, where you can turn on the Telnet. So, while Telnet Client is intended to work with the Telnet protocol, it actually is very powerful for the Transmission Control Protocol connectivity. It is an added advantage that the Telnet Client can be used to work with the Telnet protocols and can be very usefully be used to check or test the transmission control protocols connectivity. It is really because of the simplicity of the Telnet protocol itself, as it is easy to use. It basically requires establishing a TCP connection to a port and then sends and receives ASCII characters to and from the destination system through a TCP connection to a port.

This can be tested by using telnet with the following syntax: `telnet <host> <port>`. The telnet client will attempt to connect to the remote system on the specified port. If this is successful, you'll see a successful message report. If it is unsuccessful, it shows a message indicating that the connection request has been rejected or timed out. Either of these latter responses can indicate that the remote machine is not listening on that port or that the communication has been blocked. The best next step is to check the destination system and ensure that the port being expected actually has a service listening on that port. If so, the communication may be blocked by a firewall or another network device. Telnet is an old, yet very reliable communication protocol. It was originally developed as a character-oriented terminal emulation protocol used in the UNIX environment.

Today Telnet is used extensively for system administration of routers, switches, and remote servers as well as basic text communication in which graphics are not required. Although Telnet still remains a simple client/server protocol, new enhancements have been added to some products, utilizing additional local (client) processing.

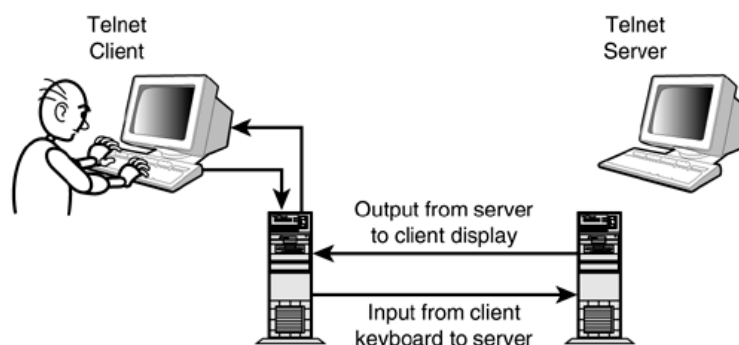


Figure 12.1 Working of Telnet

After the initial TCP handshake, the Telnet protocol performs a variety of basic housekeeping tasks known as Telnet option negotiations. These are:

- DO
- DON'T
- WILL
- WON'T

Characteristics of Telnet

Now, let us see the different characteristics one by one.

1. **Time Sharing Environment:** The first characteristic is time sharing environment. It means that telnet was designed at a time when most operating systems such as Unix were operating in timesharing environment. Telnet was first made for the Unix environment, and then seeing its strength. It was extended to variable kind of different environments, but now it is available for Linux, Windows, and Macintosh also. So, in such an environment, a large computer supports multiple users. It can have multiple users. So, initially, when users were using it, they were not using it was basically being used for the time-sharing environments. So that is its biggest strength that we have another concept or another characteristic of Telnet is logging.
2. **Logging:** It means that in Telnet environment, users are part of the system with some right to access the resources. They have some rights to the resources, which they want to access the system. The user logs into the system, he need a login ID with which he can log into the system. The user can login, in two different ways,

a) Local Log-in

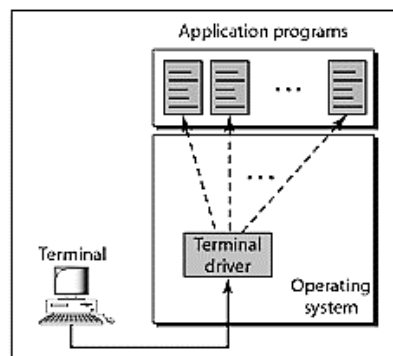


Figure 12.2 Local Log-in

As shown in Figure 12.1, there is a terminal, and inside that particular terminal user is trying to access something. There might be some application programs, as in a user's machine a user might be having a few application programs like Microsoft Office, Corel DRAW, Dream Weaver etc. So a user can have different software's inside his machine. So when he tries to log into his system, and he is trying to access any of these particular application programs and is trying to fetch some data from them. So the operating system acts as a mediator and provides it with a terminal driver which helps the user to connect with the various application programs from which he wants to fetch. So, this kind of login is called the local login, where the user is trying to log in into the local machine.

b) Remote Log-in

In case of remote login, a terminal, which is distant at one place, maybe, at one part of the country, is trying to fetch data from the Telnet server, which is at some other part of the country, and I'm trying to fetch my data from there. This kind of connection, which makes use of Internet, and with the help of which my Telnet Client connects to the Telnet server via the internet. This kind of login is called remote login.

As it can be seen in Figure 12.3, a terminal driver is being used. It does not give access to any application programs, but it is providing access to is the Telnet Client. The Telnet Client

will have different kinds of protocols like the TCP and the Internet Protocol. So, this will help me connect to the internet and via the internet, I will again be connecting with the TCP and the Internet Protocol, or in other words the Data Link and the physical layers of the Telnet server. So a user is able to connect from his client side, protocols to the server side. The pseudo terminal driver, which helps me now to connect to the application programmes, so they realise the difference I don't have a terminal driver in this case, at the server end but what I will have will have a pseudo terminal driver, which will help me connect to the different application programmes, so whatever I'm trying to fetch, whatever my Telnet Client was trying to fetch that Telnet Client via that terminal driver is making requests based upon the different layers of the TCP IP model, so it will have the TCP IP DataLink and the physical layer, making requests, and then via internet those requests are going at the different layers to the server, where the pseudo terminal driver will fetch the required application programmes and the data which are required from them, and then it will respond to you. So this is how the remote login take place. So in this case also we require some credentials, we will be requiring the login ID and the authentication rights to access the same.

3. Network Virtual Terminal
4. Embedding
5. Options

Logging

In a timesharing environment, users are part of the system with some right to access resources.

To access the system, the user logs into the system with a user id or log-in name.

a) Local Log-in

Remote Log-in

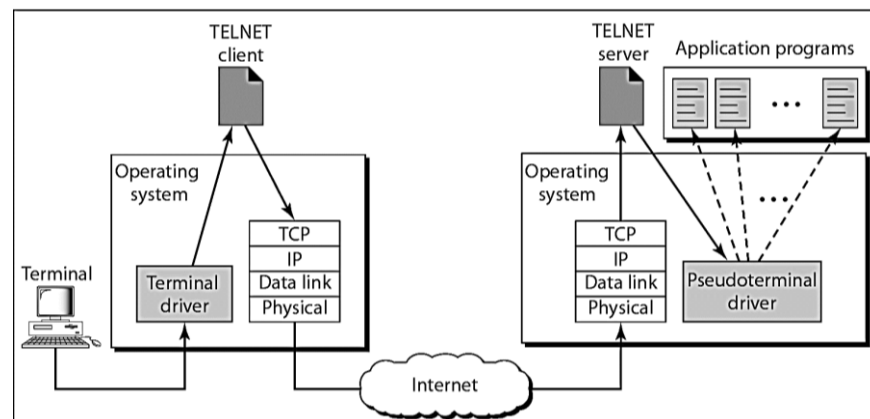


Figure 12. 3 Remote Log-in

Network Virtual Terminal

We are dealing with heterogeneous systems. If we want to access any remote computer in the world, we must first know what type of computer we will be connected to, and we must also install the specific terminal emulator used by that computer. TELNET solves this problem by defining a universal interface called the network virtual terminal (NVT) character set. Via this interface, the client TELNET translates characters (data or commands) that come from the local terminal into NVT form and delivers them to the network. The server TELNET, on the other hand, translates data and commands from NVT form into the form acceptable by the remote computer.

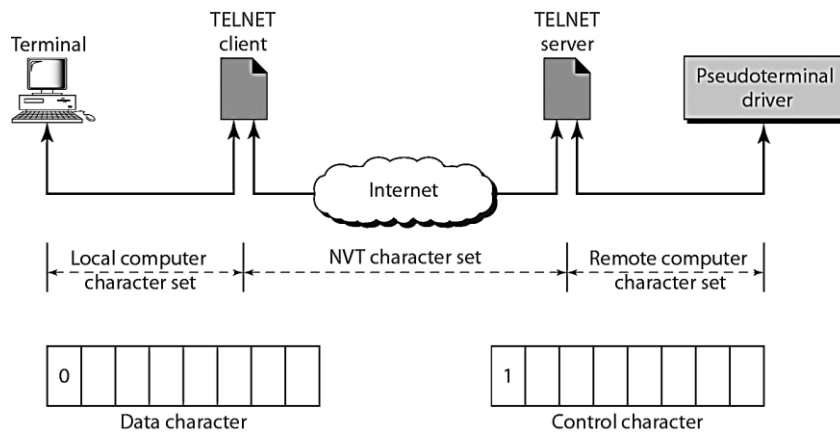


Figure 12. 4 Network Virtual Terminal

12.2 FTP

FTP stands for File transfer protocol.

FTP is a standard internet protocol provided by TCP/IP used for transmitting the files from one host to another.

It is mainly used for transferring the web page files from their creator to the computer that acts as a server for other computers on the internet.

It is also used for downloading the files to computer from other servers.

It is a way to download, upload, and transfer files from one location to another on the internet and between computer systems.

It enables computers on the internet to transfer files back and forth, and is an essential tool for those building and maintaining websites today.

Many file transfer protocol (FTP) clients are available for free to download, although most websites (and web browsers) that offer downloads already have the FTP built-in, so downloading a separate piece of software isn't always required.

Understanding File Transfer Protocol

File transfer protocol is one of many different protocols that dictate how computers behave on the internet. Other such protocols include the Hypertext Transfer Protocol (HTTP), the Internet Message Access Protocol (IMAP), and the Network Time Protocol (NTP). FTP enables computers on the internet to transfer files back and forth and is an essential tool for those building and maintaining websites today. In order to use FTP, a user must first download an FTP client (or access an FTP client through a web browser). A client is the software that will allow you to transfer files. Most web browsers come with FTP clients—possibly via a downloadable extension—that enable users to transfer files from their computer to a server and vice versa. Some users may want to use a third-party FTP client because many of them offer extra features to improve your experience.



FTP clients that are free to download include FileZilla Client, FTP Voyager, WinSCP, CoffeeCup Free FTP, and Core FTP.

Many people have used FTP before without even noticing it. If you have ever downloaded a file from a web page, chances are that you used FTP in the process.

The first step for accessing an FTP server to download a file is to log in, which may occur automatically or by manually inputting a username and password.

FTP will also require you to access an FTP server through a specific port number.

Once you have accessed the FTP server through your FTP client, you can now transfer files. Not all public FTP servers require you to sign in because some servers enable you to access them anonymously.

Depending on the FTP client you use, there will be different features available that allow you to modify the manner in which you upload and download files.

For instance, if you use the free FTP client FileZilla, the program will enable you to set bandwidth limits for files, enabling you to control the speed at which you download or upload files.

This can be helpful if you are managing multiple file transfers at once.

Other features you may want to look for in an FTP client include public key authentication, the ability to set file compression levels, or tools that enable you to search a server using file masks.

File Transfer Protocol (FTP) Example

FTP software is relatively straightforward to setup. FileZilla is a free, downloadable FTP client. Type in the address of the server you wish to access, the port, and the password for accessing the server.

FTP software is relatively straightforward to setup. FileZilla is a free, downloadable FTP client. Type in the address of the server you wish to access, the port, and the password for accessing the server.

Objectives of FTP

- It provides the sharing of files.
- It is used to encourage the use of remote computers.
- It transfers the data more reliably and efficiently.

Why FTP?

Although transferring files from one system to another is very simple and straightforward, but sometimes it can cause problems.

For example, two systems may have different file conventions. Two systems may have different ways to represent text and data. Two systems may have different directory structures.

FTP protocol overcomes these problems by establishing two connections between hosts.

One connection is used for data transfer, and another connection is used for the control connection.

Mechanism of FTP?

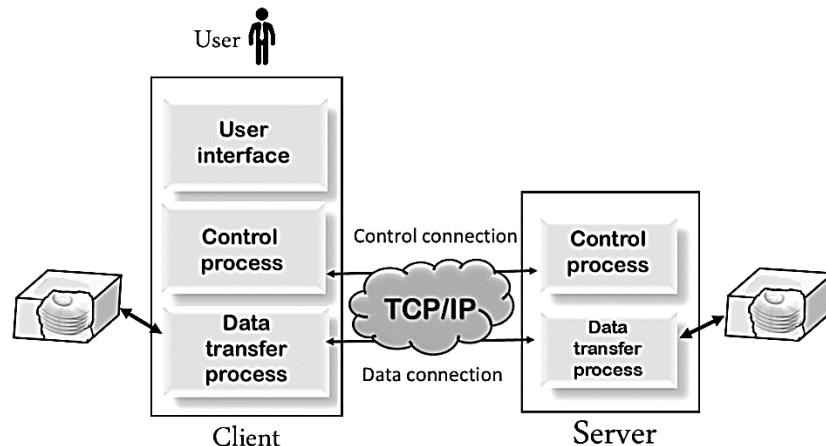


Figure 12.5 Mechanism of FTP

The above figure shows the basic model of the FTP. The FTP client has three components:

- the user interface,
- control process,
- and data transfer process.

The server has two components: the server control process and the server data transfer process.

Types of Connections in FTP?

There are two types of connections in FTP:

1. Control Connection:

The control connection uses very simple rules for communication. Through control connection, we can transfer a line of command or line of response at a time. The control connection is made between the control processes. The control connection remains connected during the entire interactive FTP session.

2. Data Connection:

The Data Connection uses very complex rules as data types may vary. The data connection is made between data transfer processes. The data connection opens when a command comes for transferring the files and closes when the file is transferred.

FTP Clients

FTP client is a program that implements a file transfer protocol which allows you to transfer files between two hosts on the internet. It allows a user to connect to a remote host and upload or download the files.

It has a set of commands that we can use to connect to a host, transfer the files between you and your host and close the connection. The FTP program is also available as a built-in component in a Web browser. This GUI based FTP client makes the file transfer very easy and also does not require to remember the FTP commands.

Advantages of FTP:

Speed: One of the biggest advantages of FTP is speed. The FTP is one of the fastest way to transfer the files from one computer to another computer. **Efficient:** It is more efficient as we do not need to complete all the operations to get the entire file.

Advantages of FTP:

Security: To access the FTP server, we need to login with the username and password. Therefore, we can say that FTP is more secure. **Back & forth movement:** FTP allows us to transfer the files back and forth. Suppose you are a manager of the company, you send some information to all the employees, and they all send information back on the same server.

Disadvantages of FTP:

- The standard requirement of the industry is that all the FTP transmissions should be encrypted. However, not all the FTP providers are equal and not all the providers offer encryption. So, we will have to look out for the FTP providers that provides encryption.
- FTP serves two operations, i.e., to send and receive large files on a network. However, the size limit of the file is 2GB that can be sent. It also doesn't allow you to run simultaneous transfers to multiple receivers.
- Passwords and file contents are sent in clear text that allows unwanted eavesdropping. So, it is quite possible that attackers can carry out the brute force attack by trying to guess the FTP password.
- It is not compatible with every system.

12.3 Domain Name System

- The Domain Name System (DNS) is a hierarchical and decentralized naming system for computers, services, or other resources connected to the Internet or a private network.
- It associates various information with domain names assigned to each of the participating entities.
- Most prominently, it translates more readily memorized domain names to the numerical IP addresses needed for locating and identifying computer services and devices with the underlying network protocols.
- By providing a worldwide, distributed directory service, the Domain Name System has been an essential component of the functionality of the Internet since 1985.
- The Domain Name System delegates the responsibility of assigning domain names and mapping those names to Internet resources by designating authoritative name servers for each domain.

- Network administrators may delegate authority over sub-domains of their allocated name space to other name servers.
- This mechanism provides distributed and fault-tolerant service and was designed to avoid a single large central database.
- The Domain Name System also specifies the technical functionality of the database service that is at its core.
- It defines the DNS protocol, a detailed specification of the data structures and data communication exchanges used in the DNS, as part of the Internet Protocol Suite.
- The Internet maintains two principal namespaces, the domain name hierarchy and the Internet Protocol (IP) address spaces.
- The Domain Name System maintains the domain name hierarchy and provides translation services between it and the address spaces.
- Internet name servers and a communication protocol implement the Domain Name System.
- A DNS name server is a server that stores the DNS records for a domain;
- A DNS name server responds with answers to queries against its database.
- The most common types of records stored in the DNS database are for Start of Authority (SOA), IP addresses (A and AAAA), SMTP mail exchangers (MX), name servers (NS), pointers for reverse DNS lookups (PTR), and domain name aliases (CNAME).
- Although not intended to be a general purpose database, DNS has been expanded over time to store records for other types of data for either automatic lookups, such as DNSSEC records, or for human queries such as responsible person (RP) records.
- As a general purpose database, the DNS has also been used in combating unsolicited email (spam) by storing a Real-time Black-hole List (RBL).
- The DNS database is traditionally stored in a structured text file, the zone file, but other database systems are common.
- The domain name system (DNS) is a naming database in which internet domain names are located and translated into internet protocol (IP) addresses.
- The domain name system maps the name people use to locate a website to the IP address that a computer uses to locate a website.
- For example, if someone types example.com into a web browser, a server behind the scenes will map that name to the corresponding IP address, something similar in structure to 121.12.12.121.
- Web browsing and most other internet activities rely on DNS to quickly provide the information necessary to connect users to remote hosts.
- DNS mapping is distributed throughout the internet in a hierarchy of authority.
- Access providers and enterprises, as well as governments, universities and other organizations, typically have their own assigned ranges of IP addresses and an assigned domain name.
- They also typically run DNS servers to manage the mapping of those names to those addresses.
- Most URLs are built around the domain name of the web server that takes client requests.

DNS in Application Layer

- DNS is a host name to IP address translation service. DNS is a distributed database implemented in a hierarchy of name servers. It is an application layer protocol for message exchange between clients and servers.

DNS Requirement

- Every host is identified by the IP address but remembering numbers is very difficult for the people and also the IP addresses are not static therefore a mapping is required to change the domain name to IP address.

- So DNS is used to convert the domain name of the websites to their numerical IP address.

DNS Domain Types

There are various kinds of DOMAIN :

- Generic domain
Domains like .com(commercial), .edu(educational), .mil(military) .org(non profit organization) .net(similar to commercial) all these are generic domain.
- Country domain
Domains like .in (india) .us .uk
- Inverse domain
If we want to know what is the domain name of the website. IP to domain name mapping.
So DNS can provide both the mapping for example to find the ip addresses of abc.org then we have to type nslookup www.abc.org.

12.4 Organization of Domain

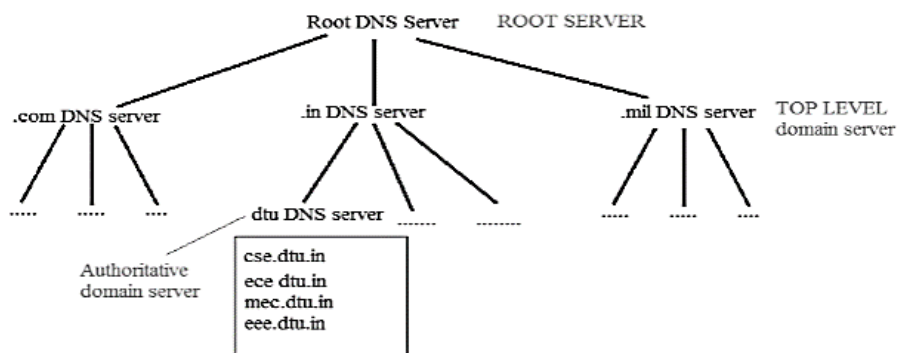


Figure 12.6 Organization of Domain

- It is very difficult to find out the ip address associated to a website because there are millions of websites and with all those websites we should be able to generate the ip address immediately,
- there should not be a lot of delay for that to happen organization of database is very important.

DNS Record

It includes

- Domain name,
- IP address
- What is the validity??
- What is the time to live ??
- and all the information related to that domain name.

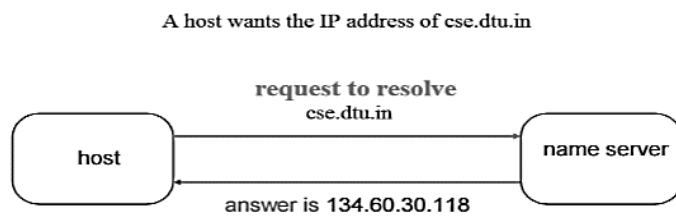
These records are stored in tree like structure.

Namespace

- Namespace is the set of possible names,
- flat or hierarchical.
- Naming system maintains a collection of bindings of names to values – given a name, a resolution mechanism returns the corresponding value –

Name Server

- It is an implementation of the resolution mechanism.
- DNS (Domain Name System) = Name service in Internet – Zone is an administrative unit, domain is a subtree.

Name to Address Resolution*Figure 12. 7 Name to Address Resolution*

- The host request the DNS name server to resolve the domain name.
- And the name server returns the IP address corresponding to that domain name to the host so that the host can future connect to that IP address.

Hierarchy of Name Servers

Let us understand the different Name Servers in the hierarchy of Name Servers

- a) Root name servers
 - b) Top level server
 - c) Authoritative name servers
- a) **Root Name Servers –**
 - It is contacted by name servers that can not resolve the name.
 - It contacts authoritative name server if name mapping is not known.
 - It then gets the mapping and return the IP address to the host.
 - b) **Top Level Server –**
 - It is responsible for com, org, edu etc and all top level country domains like uk, fr, ca, in etc.
 - They have info about authoritative domain servers and know names and IP addresses of each authoritative name server for the second level domains.
 - c) **Authoritative Name Servers –**
 - This is organization's DNS server, providing authoritative hostName to IP mapping for organization servers.
 - It can be maintained by organization or service provider.
 - In order to reach cse.dtu.in we have to ask the root DNS server, then it will point out to the top level domain server and then to authoritative domain name server which actually contains the IP address.
 - So the authoritative domain server will return the associative ip address.

Domain Name Server

The client machine sends a request to the local name server, which, if root does not find the address in its database, sends a request to the root name server, which in turn, will route the query to an intermediate or authoritative name server.

The root name server can also contain some hostName to IP address mappings.

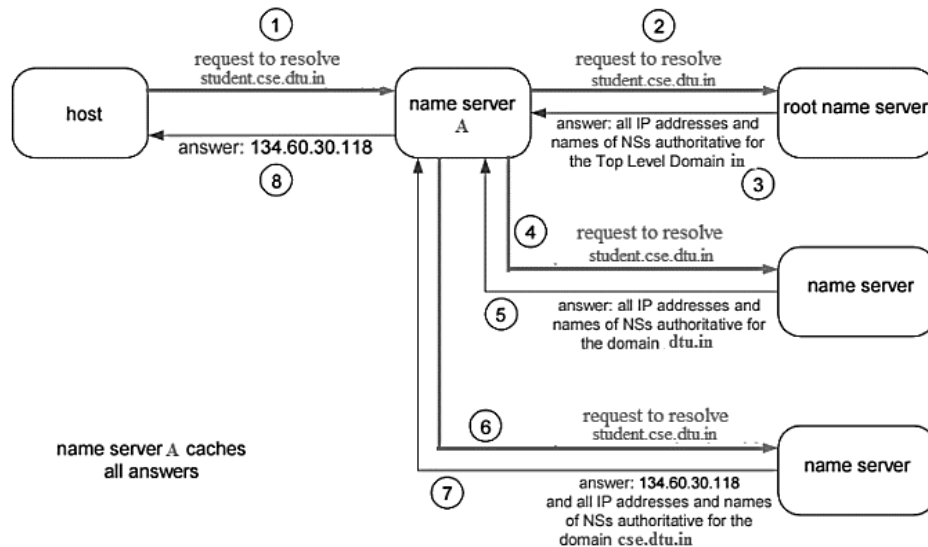


Figure 12. 8 Domain Name Server

- The intermediate name server always knows who the authoritative name server is. So finally the IP address is returned to the local name server which in turn returns the IP address to the host.

How Domain Name System Works

- DNS servers answer questions from both inside and outside their own domains.
- When a server receives a request from outside the domain for information about a name or address inside the domain, it provides the authoritative answer.
- When a server receives a request from inside its own domain for information about a name or address outside that domain, it passes the request out to another server.
- Usually, this server is one managed by its internet service provider (ISP).
- If that server does not know the answer or the authoritative source for the answer, it will reach out to the DNS servers for the top-level domain for e.g., for all of .com or .edu.
- Then, it will pass the request down to the authoritative server for the specific domain -- e.g., google.com or lpu.co.in. The answer will flow back along the same path.

Domain Name System Structure

- A domain name is made of multiple parts, called labels.
- The domain hierarchy is read from right to left with each section denoting a subdivision.
- The top-level domain is what appears after the period in the domain name.
- A few examples of top-level domains are .com, .org and .edu, but there are many others that can be used.
- Some may denote a country code or geographic location such as .in for India or .ca for Canada.
- Each label to the left denotes another subdomain to the right.
- So for example, "lpu" is a subdomain of .co.in. and "www." is a subdomain of techtarget.com.

- There can be up to 127 levels of subdomains, and each label can have up to 63 characters.
- The total domain character length can have up to 253 characters. Other rules include not starting or ending labels with hyphens and not having a fully numeric top-level domain name.
- The Internet Engineering Task Force (IETF) has specified rules considering domain names in RFC 1035, 1123, 2181 and 5892.

How DNS Increase Web Performance?

- To promote efficiency, servers can cache the answers they receive for a set amount of time. This allows them to respond more quickly the next time a request for the same lookup comes in.
- For example, if everyone in an office needs to access the same training video on a particular website on the same day, the local DNS server will ordinarily only have to resolve the name once, and then it can serve all the other requests out of its cache.
- The length of time the record is held, or the time to live, is configurable.
- Longer values decrease the load on servers, shorter values ensure the most accurate responses.

Domain Name System Security

- DNS does have a few vulnerabilities that have been discovered over time. DNS cache poisoning is one such vulnerability.
- In DNS cache poisoning, data is distributed to caching resolvers, posing as an authoritative origin server.
- The data can then present false information and can effect the time to live.
- Actual application requests can also be redirected to a malicious host network.
- An individual with malicious intent can create a dangerous website with a misleading title to try and convince users that the website they are on is real, giving the individual access to the user's information.
- By replacing a character in a domain name with a similar looking character —such as the number one “1” and a lowercase L “l,” which may look similar depending on the font—a user could be fooled into selecting a false link. This is commonly exploited with phishing attacks.
- Individuals can use DNS Security Extensions (DNSSEC) to for security, which can support cryptographically signed responses.

12.5 Simple Mail Transfer Protocol

- Various forms of one-to-one electronic messaging were used in the 1960s.
- Users communicated using systems developed for specific mainframe computers.
- As more computers were interconnected, especially in the U.S. Government's ARPANET, standards were developed to permit exchange of messages between different operating systems.
- SMTP grew out of these standards developed during the 1970s.
- SMTP traces its roots to two implementations described in 1971: the Mail Box Protocol, whose implementation has been disputed, but is discussed in RFC 196 and other RFCs, and the SNDMSG program, which, according to RFC 2235,
- Ray Tomlinson of BBN invented for TENEX computers to send mail messages across the ARPANET.
- Fewer than 50 hosts were connected to the ARPANET at this time.
- The SMTP standard was developed around the same time as Usenet, a one to many communication network with some similarities.
- Email is emerging as one of the most valuable services on the internet today.

- Most of the internet systems use SMTP as a method to transfer mail from one user to another.
- SMTP is a push protocol and is used to send the mail whereas POP (post office protocol) or IMAP (internet message access protocol) are used to retrieve those mails at the receiver's side.
- As an Internet standard, SMTP was first defined in 1982 by RFC 821, and updated in 2008 by RFC 5321 to Extended SMTP additions, which is the protocol variety in widespread use today.
- Mail servers and other message transfer agents use SMTP to send and receive mail messages.
- SMTP servers commonly use the Transmission Control Protocol on port number 25.
- User-level email clients typically use SMTP only for sending messages to a mail server for relaying, and typically submit outgoing email to the mail server on port 587 or 465 per RFC 8314.
- For retrieving messages, IMAP and POP3 are standard, but proprietary servers also often implement proprietary protocols, e.g., Exchange ActiveSync.

SMTP Fundamentals

- SMTP is an application layer protocol.
- The client who wants to send the mail opens a TCP connection to the SMTP server and then sends the mail across the connection.
- The SMTP server is always on listening mode.
- As soon as it listens for a TCP connection from any client, the SMTP process initiates a connection on that port (25).
- After successfully establishing the TCP connection the client process sends the mail instantly.

SMTP Protocol

The SMTP model is of two type :

- End-to- end Method
- Store-and- Forward Method
- The end to end model is used to communicate between different organizations whereas the store and forward method are used within an organization.
- A SMTP client who wants to send the mail will contact the destination's host SMTP directly in order to send the mail to the destination.
- The SMTP server will keep the mail to itself until it is successfully copied to the receiver's SMTP.
- The client SMTP is the one which initiates the session let us call it as the client- SMTP and the server SMTP is the one which responds to the session request and let us call it as receiver- SMTP.
- The client- SMTP will start the session and the receiver-SMTP will respond to the request.

Model of SMTP System

- In the SMTP model user deals with the user agent (UA) for example Microsoft Outlook, Netscape, Mozilla, etc.
- In order to exchange the mail using TCP, MTA is used.
- The users sending the mail do not have to deal with the MTA it is the responsibility of the system admin to set up the local MTA.
- The MTA maintains a small queue of mails so that it can schedule repeat delivery of mail in case the receiver is not available.

- The MTA delivers the mail to the mailboxes and the information can later be downloaded by the user agents.

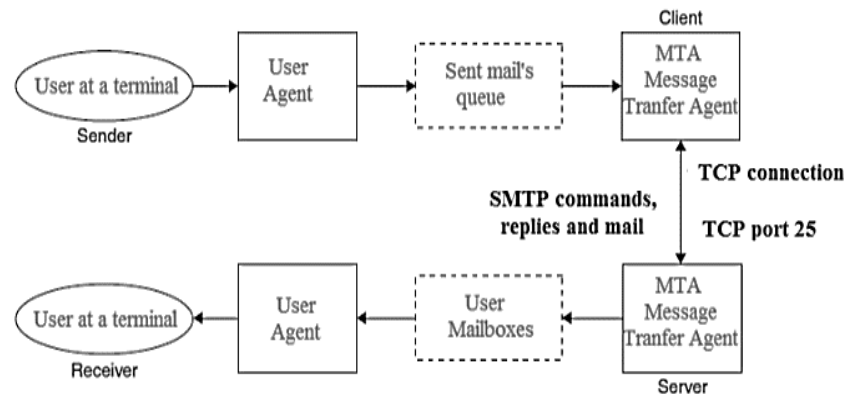


Figure 12. 9 Model of SMTP System

Both the SMTP-client and SMTP-server should have two components:

- User agent (UA)
- Local MTA

Communication Between Sender and Receiver

- The senders, user agent prepare the message and send it to the MTA.
- The MTA functioning is to transfer the mail across the network to the receivers MTA.
- To send mail, a system must have the client MTA, and to receive mail, a system must have a server MTA.

Sending Email:

- Mail is sent by a series of request and response messages between the client and a server.
- The message which is sent across consists of a header and the body.
- A null line is used to terminate the mail header.
- Everything which is after the null line is considered as the body of the message which is a sequence of ASCII characters.
- The message body contains the actual information read by the receipt.

Receiving Email:

- The user agent at the server-side checks the mailboxes at a particular time of intervals.
- If any information is received it informs the user about the mail.
- When the user tries to read the mail it displays a list of mails with a short description of each mail in the mailbox.
- By selecting any of the mail user can view its contents on the terminal.

Some SMTP Commands:

- HELO – Identifies the client to the server, fully qualified domain name, only sent once per session
- MAIL – Initiate a message transfer, fully qualified domain of originator
- RCPT – Follows MAIL, identifies an addressee, typically the fully qualified name of the addressee and for multiple addressees use one RCPT for each addressee
- DATA – send data line by line

12.6 POP Protocol

- The POP protocol stands for Post Office Protocol.

- As we know that SMTP is used as a message transfer agent.
- When the message is sent, then SMTP is used to deliver the message from the client to the server and then to the recipient server.
- But the message is sent from the recipient server to the actual server with the help of the Message Access Agent.
- The Message Access Agent contains two types of protocols, i.e., POP3 and IMAP.

How is Mail Transmitted?

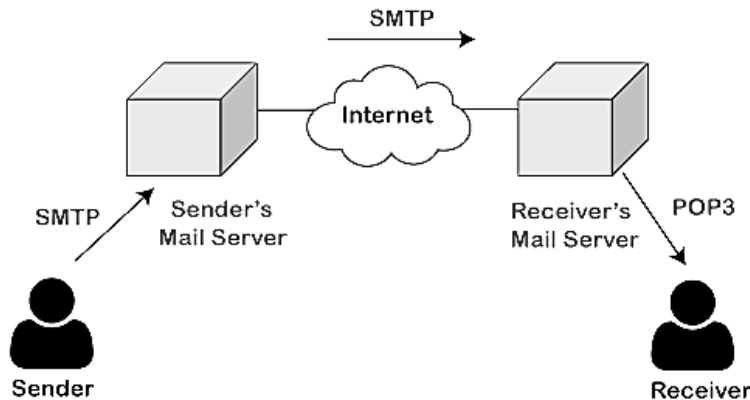


Figure 12. 10 How is Mail Transmitted

- Suppose sender wants to send the mail to receiver. First mail is transmitted to the sender's mail server.
- Then, the mail is transmitted from the sender's mail server to the receiver's mail server over the internet.
- On receiving the mail at the receiver's mail server, the mail is then sent to the user.
- The whole process is done with the help of Email protocols.
- The transmission of mail from the sender to the sender's mail server and then to the receiver's mail server is done with the help of the SMTP protocol.
- At the receiver's mail server, the POP or IMAP protocol takes the data and transmits to the actual user.
- Since SMTP is a push protocol so it pushes the message from the client to the server.
- As we can observe in the above figure that SMTP pushes the message from the client to the recipient's mail server.
- The third stage of email communication requires a pull protocol, and POP is a pull protocol.
- When the mail is transmitted from the recipient mail server to the client which means that the client is pulling the mail from the server.
- The POP3 is a simple protocol and having very limited functionalities. In the case of the POP3 protocol, the POP3 client is installed on the recipient system while the POP3 server is installed on the recipient's mail server.

History of POP3 Protocol

- The first version of post office protocol was first introduced in 1984 as RFC 918 by the internet engineering task force.
- The developers developed a simple and effective email protocol known as the POP3 protocol, which is used for retrieving the emails from the server.
- This provides the facility for accessing the mails offline rather than accessing the mailbox offline.

- In 1985, the post office protocol version 2 was introduced in RFC 937, but it was replaced with the post office protocol version 3 in 1988 with the publication of RFC 1081.
- Then, POP3 was revised for the next 10 years before it was published. Once it was refined completely, it got published on 1996.
- Although the POP3 protocol has undergone various enhancements, the developers maintained a basic principle that it follows a three-stage process at the time of mail retrieval between the client and the server.
- They tried to make this protocol very simple, and this simplicity makes this protocol very popular today.

Let's understand the working of the POP3 protocol.

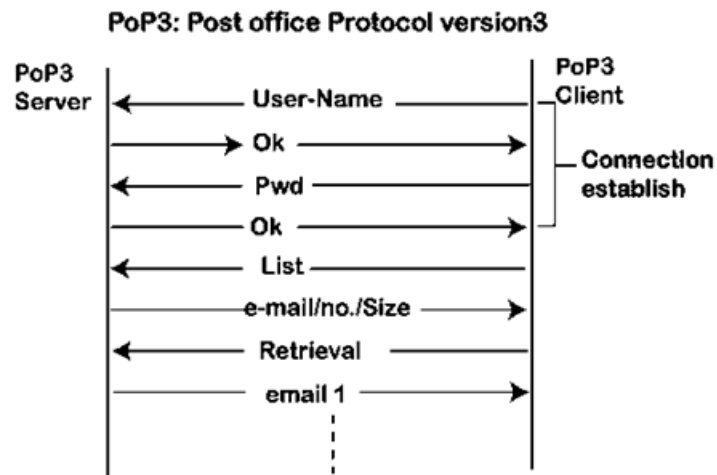


Figure 12. 11 POP3 Protocol

- To establish the connection between the POP3 server and the POP3 client, the POP3 server asks for the user name to the POP3 client.
- If the username is found in the POP3 server, then it sends the ok message.
- It then asks for the password from the POP3 client; then the POP3 client sends the password to the POP3 server.
- If the password is matched, then the POP3 server sends the OK message, and the connection gets established.
- After the establishment of a connection, the client can see the list of mails on the POP3 mail server.
- In the list of mails, the user will get the email numbers and sizes from the server.
- Out of this list, the user can start the retrieval of mail.
- Once the client retrieves all the emails from the server, all the emails from the server are deleted.
- Therefore, we can say that the emails are restricted to a particular machine, so it would not be possible to access the same mails on another machine.
- This situation can be overcome by configuring the email settings to leave a copy of mail on the mail server.

Advantages of POP3 Protocol

- It allows the users to read the email offline.
 - It requires an internet connection only at the time of downloading emails from the server.

- Once the mails are downloaded from the server, then all the downloaded mails reside on our PC or hard disk of our computer, which can be accessed without the internet.

Therefore, we can say that the POP3 protocol does not require permanent internet connectivity.

It provides easy and fast access to the emails as they are already stored on our PC.

There is no limit on the size of the email which we receive or send.

It requires less server storage space as all the mails are stored on the local machine.

There is maximum size on the mailbox, but it is limited by the size of the hard disk.

It is a simple protocol so it is one of the most popular protocols used today.

It is easy to configure and use.

Disadvantages of POP3 Protocol

- If the emails are downloaded from the server, then all the mails are deleted from the server by default. So, mails cannot be accessed from other machines unless they are configured to leave a copy of the mail on the server.
- Transferring the mail folder from the local machine to another machine can be difficult.
- Since all the attachments are stored on your local machine, there is a high risk of a virus attack if the virus scanner does not scan them. The virus attack can harm the computer.
- The email folder which is downloaded from the mail server can also become corrupted.
- The mails are stored on the local machine, so anyone who sits on your machine can access the email folder.

12.7 Internet Message Access Protocol (IMAP)

- Internet Message Access Protocol (IMAP) is an application layer protocol that operates as a contract for receiving emails from the mail server.
- It was designed by Mark Crispin in 1986 as a remote access mailbox protocol, the current version of IMAP is IMAP4.
- Internet Message Access Protocol (IMAP)
- It is used as the most commonly used protocol for retrieving emails.
- This term is also known as Internet mail access protocol, Interactive mail access protocol, and Interim mail access protocol.

Features of IMAP

- It is capable of managing multiple mailboxes and organizing them into various categories.
- Provides adding of message flags to keep track of which messages are being seen.
- It is capable of deciding whether to retrieve email from a mail server before downloading.
- It makes it easy to download media when multiple files are attached.

Working of IMAP

- IMAP follows Client-server Architecture and is the most commonly used email protocol.
- It is a combination of client and server process running on other computers that are connected through a network.
- This protocol resides over the TCP/IP protocol for communication.
- Once the communication is set up the server listens on port 143 by default which is non-encrypted.
- For the secure encrypted communication port, 993 is used.

Architecture of IMAP

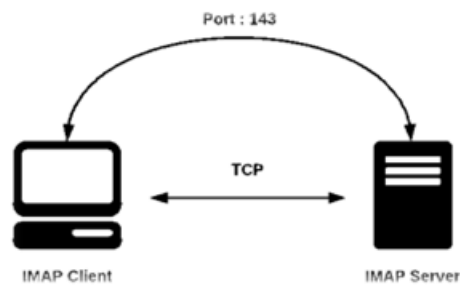


Figure 12. 12 Architecture of IMAP

Advantages of IMAP

- It offers synchronization across all the maintained sessions by the user.
- It provides security over POP3 protocol as the email only exists on the IMAP server.
- Users have remote access to all the contents.
- It offers easy migration between the devices as it is synchronized by a centralized server.
- There is no need to physically allocate any storage to save contents.

Disadvantages of IMAP

- IMAP is complex to maintain.
- Emails of the user are only available when there is an internet connection.
- It is slower to load messages.
- Some emails don't support IMAP which makes it difficult to manage.
- Many browser-based solutions are unavailable due to not support of IMAP.

Internet Message Access Protocol

- Since IP does not have a built-in mechanism for sending error and control messages. It depends on Internet Control Message Protocol(ICMP) to provide an error control.
- It is used for reporting errors and management queries.
- It is a supporting protocol and used by network devices like routers for sending the error messages and operations information.
- e.g. the requested service is not available or that a host or router could not be reached.
- IMAP - Source Quench message
- Source quench message is request to decrease traffic rate for messages sending to the host(destination).
- Or we can say, when receiving host detects that rate of sending packets (traffic rate) to it is too fast it sends the source quench message to the source to slow the pace down so that no packet can be lost.

IMAP - Source Quench Message

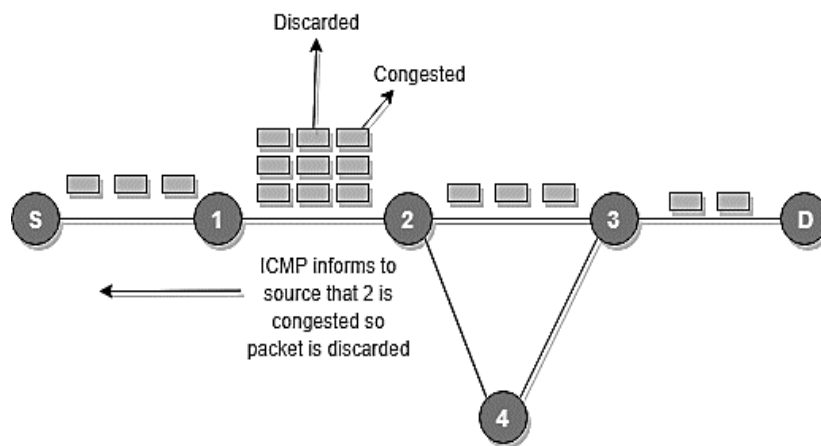


Figure 12. 13 IMAP - Source Quench message

- ICMP will take source IP from the discarded packet and informs to source by sending source quench message.
- Then source will reduce the speed of transmission so that router will free for congestion.

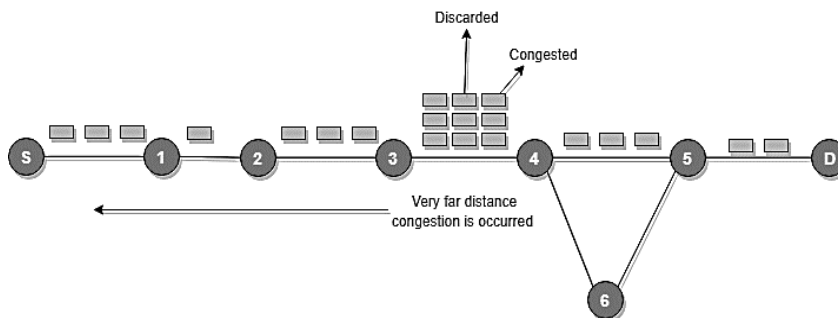


Figure 12. 14 IMAP - Source Quench message (Reduced Speed of Transmission)

- When the congestion router is far away from the source the ICMP will send hop by hop source quench message so that every router will reduce the speed of transmission.

IMAP - Parameter Problem

- Whenever packets come to the router then calculated header checksum should be equal to received header checksum then only packet is accepted by the router.

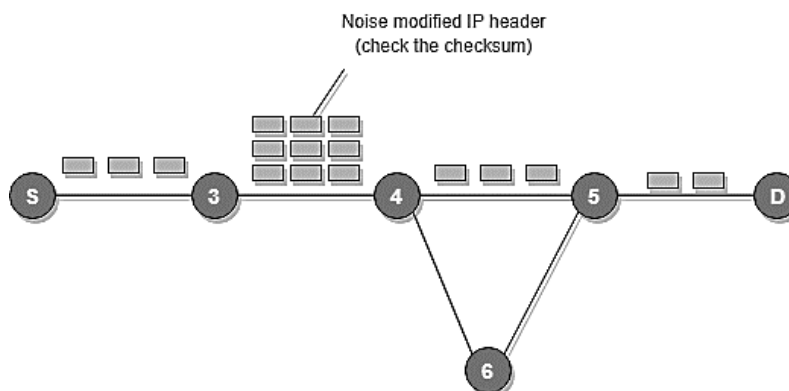


Figure 12. 15 IMAP - Parameter Problem

- If there is mismatch packet will be dropped by the router.

- ICMP will take the source IP from the discarded packet and informs to source by sending parameter problem message.

IMAP – Time Exceeded Message

- When some fragments are lost in a network then the holding fragment by the router will be dropped,
- then ICMP will take source IP from discarded packet and informs to the source, of discarded datagram due to time to live field reaches to zero, by sending time exceeded message.

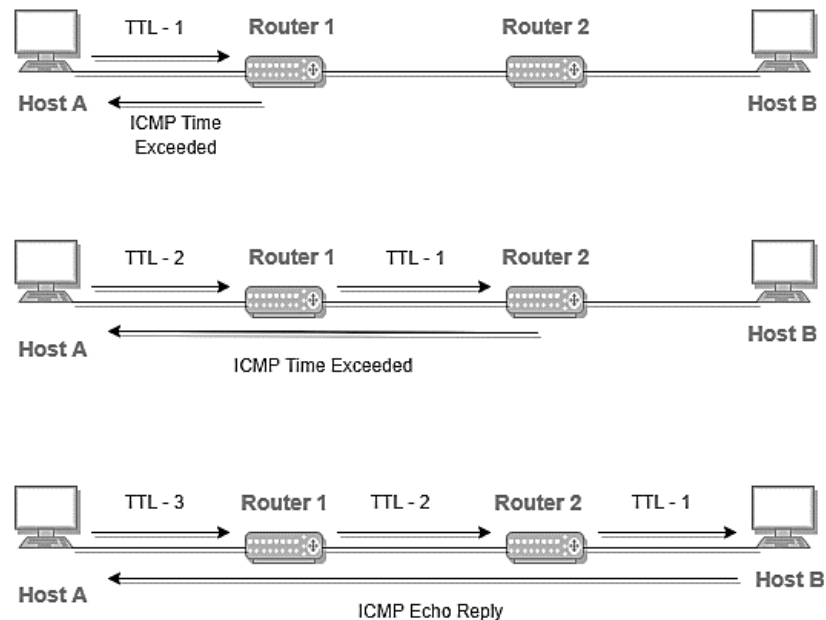


Figure 12. 16 IMAP – Time Exceeded Message

IMAP – Destination Un-Reachable

- Destination unreachable is generated by the host or its inbound gateway to inform the client that the destination is unreachable for some reason.
- There is no necessary condition that only router give the ICMP error message some time destination host send ICMP error message when any type of failure (link failure, hardware failure, port failure etc.) happen in the network.

IMAP – Time Exceeded Message

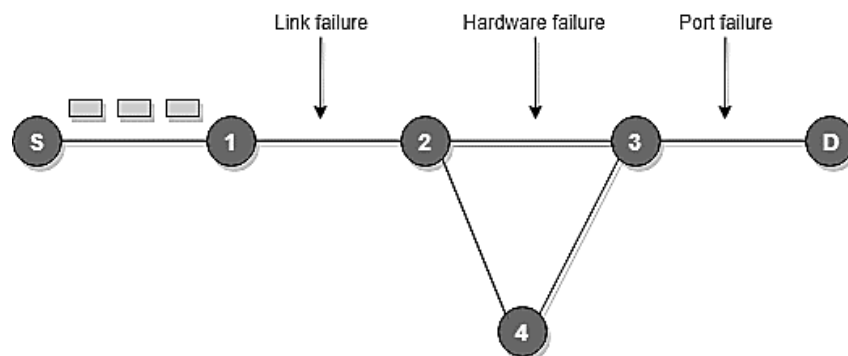


Figure 12. 17 IMAP – Time Exceeded Message

IMAP – Redirection Message

- Redirect requests data packets be sent on an alternate route.
- The message informs to a host to update its routing information (to send packets on an alternate route).

- Ex. If host tries to send data through a router R1 and R1 sends data on a router R2 and there is a direct way from host to R2. Then R1 will send a redirect message to inform the host that there is a best way to the destination directly through R2 available. The host then sends data packets for the destination directly to R2.
- The router R2 will send the original datagram to the intended destination.
- But if datagram contains routing information then this message will not be sent even if a better route is available as redirects should only be sent by gateways and should not be sent by Internet hosts.

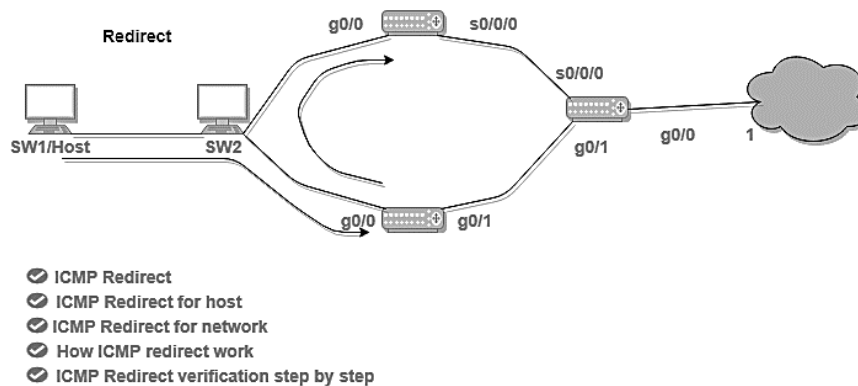


Figure 12. 18 ICMP – Redirect Verification Message

- Whenever a packet is forwarded in a wrong direction later it is re-directed in a current direction then ICMP will send re-directed message.

Summary

Having discussed the various transmission medias and the different type of signals that they deal with, a proper transmission mode has to be chosen for the type of communication we are undergoing. To ensure the effective and error-free transmission different essential network performance metrics must be considered to make necessary changes in the network. The different transmission impairments need to be understood by an organisation and should be effectively managed to ensure effective data communication

Keywords

Archive: A computer site advertises and stores a large amount of public domain, shareware software and documentation.

Broadcast Networks: They have a single communication channel, which is shared by all the computers on the network and therefore, any message transmitted by a computer on the network is received by all the computers connected to the channel.

Bit Length: It is the distance one bit occupies on the transmission medium.

Baud Rate is the number of signal unit transmitted per second. It is always less than or equal to bit rate.

Distortion: It means changes in the form or shape of the signal. This is generally seen in composite signals made up with different frequencies. Each frequency component has its own propagation speed travelling through a medium. And that's why it delays in arriving at the final destination

Self-Assessment

Fill in the blanks

1.refers to whether the connections between the nodes on your network are working properly.
2.measures your network's actual data transmission rate, which can vary wildly through different areas of your network.

3.is the delay that happens between a node or device requesting data and when that data is finished being delivered.
4.examines how many data packets are dropped during data transmissions on your network.
5.rate lets your enterprise know how often packets are being dropped, which is an indication of congestion on your network.

State whether the following is true or false.

6. Bandwidth is the minimum data transmission rate possible on a network.
7. De facto standards are the standards that are followed with a formal plan or approval by any organization.
8. De jure standards are the standards which have been adopted through legislation by any officially recognized standards organization.
9. The World Wide Web Consortium (W3C) is the main international standards organization for World Wide Web which was founded and headed by Tim Berners-Lee.
10. SNR measures the quality of a system that indicates the strength of the signal wrt the noise power in the system.

Review Questions

1. Underline the key differences between Bit Rate and Baud Rate. Also elaborate on Bit length and Bit Interval.
2. Signals can be classified based on different parameters. Elaborate the different classification categories.
3. Explain the different factors that can affect a Network Performance. Explain the metrics that are essential for any businesses to consider.
4. Explain how the imperfections in the transmission medias causes signal impairments. Explain the various types of impairments.
5. Compare and contrast the various types of transmission modes. Take suitable examples to explain.

Answers: Self Assessment

- | | |
|-------------------|----------------|
| 1. Connectivity | 2. Throughput |
| 3. Latency | 4. Packet loss |
| 5. Retransmission | 6. False |
| 7. False | 8. True |
| 9. True | 10. True |

Further Readings



Andrew S. Tanenbaum, *Computer Networks*, Prentice Hall

Behrouz A. Forouzan and Sophia Chung Fegan, *Data Communications and Networking*, McGraw-Hill Companies

Burton, Bill, *Remote Access for Cisco Networks*, McGraw-Hill Osborne Media McGraw-Hill Osborne Media

Rajneesh Agrawal and Bharat Bhushan Tiwari, *Computer Networks and Internet*, Vikas Publication

Unit 13 : Introduction to Computer Networks

CONTENTS

Objectives

Introduction

13.1 Internet

13.2 Basics of Internet

13.3 URI

13.4 Who Governs the Internet?

13.5 World Wide Web (WWW)

Summary

Keywords

Self-Assessment

Review Questions

Further Readings

Objectives

After this lecture, you would be able to:

- learn the basics of Internet
- understand why Internet is called a Network
- understand the working of Internet
- learn the various advantages and disadvantages
- learn the basics of HTTP and explore its characteristics
- understand how the HTTP protocol works
- explore the various advantages and disadvantages of HTTP
- understand the differences between HTTP and HTTPS
- learn about the basics of WWW
- understand its System Architecture
- understand the Working of WWW
- learn the features and components of WWW
- Understand the Virtual Private Networks
- Learn the different types of VPN
- Understand the various security mechanisms
- Learn the Life Cycle phases of IPSec Tunnel in VPN

Introduction

Congestion is any stage, into which we aren't able to carry on the normal communication. As you can see on the screen. This is a typical example of a traffic congestion, where there is converging traffic. And at any particular point of time there is a choking situation in the traffic. So the traffic is not having a smooth flow and this is what we mean by congestion. So, it is the reduced quality of service that occurs when a network node or link is carrying more data than it can handle. So when will a traffic overflow occur. A traffic overflow would only occur when a particular road is carrying traffic more than its own capacity. If it is carrying traffic, more than the capacity. It is only under such

circumstances that an overflow or a congestion situation can occur. Now typical effects of this congestion include queuing delay. That means you will have long waiting in the queue packet loss packets might be lost during transmission or blocking of new connections, so new connections will be blocked, they will not be able to traverse on that particular network. So, this is about congestion. So, what is a consequence of congestion, well, a consequence of congestion, is that an incremental increase in offered load, leads, either only to a small increase, or even a decrease in the network throughput. So, the performance of the network depreciates. Now, let us discuss the basic requirements, or the quality of service specifications. Now, quality of service requirements can be classified or specified as belonging to the four factors. So, quality of service can be specified as delay. Delay variation which I've already told you it is called jitter. It could be on the basis of throughput, that that is the output that I get after a particular time period, and the error rate which is there, and there are two types of quality of service solutions, which I, which are available for us. One is the stateless solution. Another is the stateful solution. So, what are these, let us see one by one. In stateless solution routers maintain no fine-grained state about traffic. So one positive factor of it is that it is scalable, and it is robust, but it has weak services, as there is no guarantee about the kind of delay or performance in a particular application, which we have to encounter, talking about the integrated services, or the in itself, what exactly do we mean by that. Let us now see an architecture for providing quality of service guarantees in IP networks for individual application sessions. So, the architecture, which is to provide quality of service guarantee, and in, in the IP network because it is the Internet Protocol networks that we use so individual application sessions need to take care of this. It relies on the resource reservation and routers need to maintain state information of allocated resources.

13.1 Internet

Internet is a global network that connects billions of computers with each other and to the world wide web. It uses standard internet protocol suite that is TCP/IP to connect billions of computer users worldwide. It is set up by using cables such as optical fibers and other wireless and networking technologies. So, at present Internet is the fastest means of sending and exchanging information and data between computers across the world. It is believed that internet was developed by Defense Advanced Research Projects Agency (DARPA), and it was first connected in the year 1969.

Why is the Internet Called a Network?

Now, the question is why Internet is called a network. Well, to answer this question we can see that internet is called a network as it creates a network by connecting computers and servers across the world using routers, switches, telephone lines, and other communication devices and channels. So, it can be considered a global network of physical cables such as copper telephone wires, cable TV cables are the optical fibers, and many more options. Furthermore, even wireless connections like 3G, 4G and WiFi, make use of these cables to access the internet. Internet is different from the world wide web as the World Wide Web is a network of computers and servers, created by connecting them through the network. So, internet is the backbone of the web, as it provides the technical infrastructure to establish the World Wide Web, and acts as a medium to transmit information from one computer to the other. It uses web browsers to display the information on the client, which it fetches from the web servers. So, internet is not owned by a single person or an organisation entirely. It is a concept based upon physical infrastructure that connects networks with other networks to create a global network of billions of computers. As of August 12 2016, there were more than 300 crores of Internet users across the globe, which has almost doubled up by now.

13.2 Basics of Internet

Now let us understand some basics of internet, we can say that the internet works with the help of clients and servers. So, a device such as a laptop which is connected to the internet is called a client, not a server, as it is not directly connected to the internet. However, it is indirectly connected to the internet through an Internet Service Provider or an ISP and is identified by an IP address.

Just like we have an address for your home, that uniquely identifies our home, an IP address, acts as a shipping address for your device. So the IP address is provided by your internet service provider, and you can see what IP address your ISP has given to your system. So, a server is a large computer that stores, websites, it also has an IP address. Now, a place where a large number of servers are stored is called a Data Centre. The server accepts requests sent by the client through a browser over a network or internet and responds accordingly to access the internet we need a domain name, which represents an IP address number that has been assigned a domain name. For example, we might have youtube.com facebook.com paypal.com they are used to represent the IP addresses. Well, it is very hard to remember the numeric IP addresses, but remembering these URLs are pretty easy. So domain

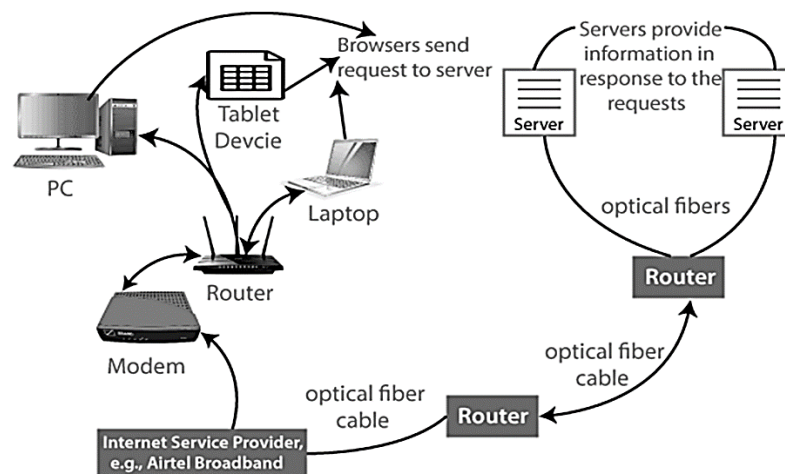
names are created as it is difficult for a person to remember a long string of numbers like 172.16.16.1. So, however Internet does not understand the domain name, what it understands is the IP address. So when we enter the domain name in the browser. In the search bar, the internet has to get the IP address for the domain name from a huge phonebook, which is generally referred to as the domain name server. Now, let us understand this with an example. Say if you have a person's name and you want to find his phone number in a phone book by searching his name. So what do we do the Internet uses the domain name server. In the same way, to find the IP address of the domain name, the domain name servers are managed by Internet service providers or similar organizations. Now, we can understand this with the help of the current diagram. As you can see in the diagram, we have a personal computer, which is being connected to another computer via a modem. We might be making use of routers to connect this PC to another laptop, which might be being connected with the help of internet. So what is governing the modem and who is connecting the modem to the net. Well, it is the internet service providers for example we might have broadband connections from the different service providers, and these service providers might be providing services with the help of maybe a optical fibre cable or a twisted pair cable now, that depends upon the service providers. So make use of a router, and this router will be doing the address resolution path for you. So if you are entering www.google.com Your computer needs to know what IP address are you looking for. To do that, you will be moving to a DNS server on the DNS server, there is a table, into which there are corresponding entries for the IP addresses corresponding to the URLs that you are entering, so corresponding to www.google.com we will be having an entry of an IP address in the DNS server that IP address is reflected back to your computer, and that is way you reach your intended website very easily. So when you turn on your computer and type a domain name in the browser's search bar, your browser sends a request to the DNS server to get the corresponding IP address. So after getting the IP address, the browser forwards the request to the respective server. Once the server gets the request to provide information about a particular website, the data starts flowing. The data is transferred through the optical fibre cables in digital format, or in the form of light pulses. So as the servers are placed at distant places that data may have to travel 1000s of miles through optical fibre cables to reach your computer. The optical fibre is connected to a router, which converts the light signals into electrical signals. These electrical signals are transmitted to your laptop using an Ethernet cable. Thus, you receive the desired information through the internet, which is actually a cable that connects you with the server. Further, if you are using a wireless internet using Wi Fi, or mobile data, the signal from the optical cables are first sent to a cell tower. And from there, it reaches your cell phone in the form of electromagnetic waves. Now let us understand who manages the internet. Well, the internet is managed by the Internet Corporation for assigned names and numbers, which is located in USA. It manages the IP addresses, assignment domain name registrations, etc. So the data transfer is very fast on the internet. The moment you press enter, you get the information from a server located 1000s of miles away from you. So the reason for this speed is that data is sent in binary forms, which is zeros and ones and these zeros and ones are divided into small pieces called packets, which can be sent to high speed. Now there are different advantages and disadvantages of internet. First let us look at the various advantages first basic advantage of the internet is instant messaging, that means you can send message or communicate to anyone using internet such as email voice chat, video conferencing, etc. Also we can get directions using GPS technology, you can get directions to almost every place in your city or country. You can find restaurants, malls or other services, near your location. Another advantage of the internet is online shopping. While it allows you to shop online such as you can be buying clothes, shoes, books, movies, or you can buy railway tickets or flight tickets, and many more things. Internet can also be used for paying the bills, you can pay your bills online, such as your electricity bills, your gas bills, or your college fees. Another advantage of the internet does online banking, it allows you to use internet banking in which you can check your balance, receive or transfer money or get a statement request, or a chequebook. Whatever you want to do with your banking transactions. Another advantage of the internet would be the online selling, you can sell your

products or services online. It helps you reach your customers and thus increase your sales and your profits. Another advantage of the internet is what we witness during the COVID period, that is work from home, in case we need to work from home you can do it by using your system with an internet access. Today many companies allow their employees to work from home. Last but not the least the biggest advantage of internet is entertainment, so you can listen to you online music, watch videos or movies or play online games. Well, internet is an integral part of life, and we can think of our life without internet. Another advantage of the internet is cloud computing. It enables you to connect your computers and internet enabled devices to cloud services such as the cloud storage cloud computing, etc. We can use internet for career building also, you can search for jobs online on different job portals and send your CV through emails, if required. Now, let us see the various

differences between internet and the web. Well, internet is the network of networks, and the network allows you to exchange the data between two or more computers, whereas the web is a way to access information through the Internet where internet is known as a network of networks web is a model for sharing information using internet. Internet is a way of transporting information between devices, whereas web is the protocol useful web is HTTP. So, when you are using web, the protocol or the set of rules is called Hypertext Transfer Protocol which we will be covering up in the next session. The web is accessed by the web browser. Last but not the least, we need to understand what we mean by a uniform resource identifier. Well, a URI can be a name or location or both of the online resources where as a URL is just the locator URLs are a subsets of the URI is a URL is a human readable text that was designed to replace the numbers or the IP addresses that companies use to communicate with the servers, a URL consists of a protocol or domain name, and a path, which includes this specific sub folder structure, where in a page, you can understand it, with the one which is written on the screen, you can see, first we have the protocol, which could be either HTTP or HTTPS. Next we have the website name. It could be any website name like you could be going to Google, or you could be going to. Yeah. Next we have the top level domain now top level domain could be .com .edu .in, for instance in the domain, if you're writing down. lpu.co.in a .in is a top level domain. And last, you mentioned the path, which is this path to the specific folders or the sub folders that there are in the website for example if you're trying to go to a login page and your link says www.lpu.co.in forward slash login dot HTML. That is a subfolder to which you are moving to. Now, a million dollar question which everybody is interested in knowing who

governs the internet. Well internet is not governed, and has a single authority figure. The ultimate authority for where the internet is going rests with the Internet Society, or the ISOC. ISOC is a voluntary membership organisation whose purpose is to promote global information exchanges, through internet technology. ISOC appoints the Internet Architecture Board. Now they meet regularly to review standards and allocate resources like addresses the Internet Engineering Task Force is another voluntary organisation that meets regularly to discuss operational and technical problems. Well that's all for now. See you again in yet another session on computer networks, until then, goodbye, and take very good care of yourselves. thank you so much.

How Internet Works



13.1 How Internet Works

When you turn on your computer and type a domain name in the browser search bar, your browser sends a request to the DNS server to get the corresponding IP address. After getting the IP address, the browser forwards the request to the respective server. Once the server gets the request to provide information about a particular website, the data starts flowing. The data is transferred through the optical fiber cables in digital format or in the form of light pulses. As the servers are placed at distant places, the data may have to travel thousands of miles through optical fiber cable to reach your computer. The optical fiber is connected to a router, which converts the light signals into electrical signals. These electrical signals are transmitted to your laptop using an Ethernet cable.

Thus, you receive the desired information through the internet, which is actually a cable that connects you with the server.

Furthermore, if you are using wireless internet using WiFi or mobile data, the signals from the optical cable are first sent to a cell tower and from where it reaches to your cell phone in the form of electromagnetic waves.

The internet is managed by ICANN (Internet Corporation for Assigned Names and Numbers) located in the USA. It manages IP addresses assignment, domain name registration, etc.

The data transfer is very fast on the internet. The moment you press enter you get the information from a server located thousands of miles away from you.

The reason for this speed is that the data is sent in the binary form (0, 1), and these zeros and ones are divided into small pieces called packets, which can be sent at high speed.

Advantages of the Internet:

Instant Messaging: You can send messages or communicate to anyone using internet, such as email, voice chat, video conferencing, etc.

Get directions: Using GPS technology, you can get directions to almost every place in a city, country, etc. You can find restaurants, malls, or any other service near your location.

Online Shopping: It allows you to shop online such as you can be clothes, shoes, book movie tickets, railway tickets, flight tickets, and more.

Pay Bills: You can pay your bills online, such as electricity bills, gas bills, college fees, etc.

Online Banking: It allows you to use internet banking in which you can check your balance, receive or transfer money, get a statement, request cheque-book, etc.

Online Selling: You can sell your products or services online. It helps you reach more customers and thus increases your sales and profit.

Work from Home: In case you need to work from home, you can do it using a system with internet access. Today, many companies allow their employees to work from home.

Entertainment: You can listen to online music, watch videos or movies, play online games.

Cloud computing: It enables you to connect your computers and internet-enabled devices to cloud services such as cloud storage, cloud computing, etc.

Career building: You can search for jobs online on different job portals and send you CV through email if required.

Difference between Internet and Web

Table 13.1 Difference between Internet and Web

Internet	Web
The Internet is the network of networks and the network allows to exchange of the data between two or more computers.	The Web is a way to access Information through the Internet.
It is also known as Network of Networks.	The Web is a model for sharing information using Internet.
The Internet is a way of transporting information between devices.	The protocol used by the web is Http.
	The Web is accessed by the Web Browser.

13.3 URI

URI stands for 'Uniform Resource Identifier'.

A URI can be a name, locator, or both for an online resource whereas a URL is just the locator.

URLs are a subset of URIs.

A URL is human-readable text that was designed to replace the numbers (IP addresses) that computers use to communicate with servers.

A URL consists of a protocol, domain name, and path (which includes the specific subfolder structure where a page is located) like-

protocol://WebSiteName.topLevelDomain/path

Protocol – Http or Https.

WebSiteName – yahoo, google etc.

topLevelDomain- .com, .edu, .in etc.

path- specific folders/subfolders that are on a website.

13.4 Who Governs the Internet?

The Internet is not governed and has no single authority figure. The ultimate authority for where the Internet is going rests with the Internet Society, or ISOC.

ISOC is a voluntary membership organization whose purpose is to promote global information exchange through Internet technology.

ISOC appoints the IAB- Internet Architecture Board. They meet regularly to review standards and allocate resources, like addresses.

IETF- Internet Engineering Task Force. Another volunteer organization that meets regularly to discuss operational and technical problems.

HTTP

- HTTP stands for HyperText Transfer Protocol. It was invented by Tim Berner.
- HyperText is the type of text which is specially coded with the help of some standard coding language called as HyperText Markup Language (HTML).
- HTTP/2 is latest version of HTTP, which was published on May 2015.
- The protocols that are used to transfer hypertext between two computers is known as HyperText Transfer Protocol.
- HTTP provides standard between a web browser and web server to establish communication.
- It is set of rules for transferring data from one computer to another.
- Data such as text, images, and other multimedia files are shared on the World Wide Web.
- Whenever a web user opens their web browser, user will indirectly use HTTP.
- It is an application protocol which is used for distributed, collaborative, hypermedia information systems.

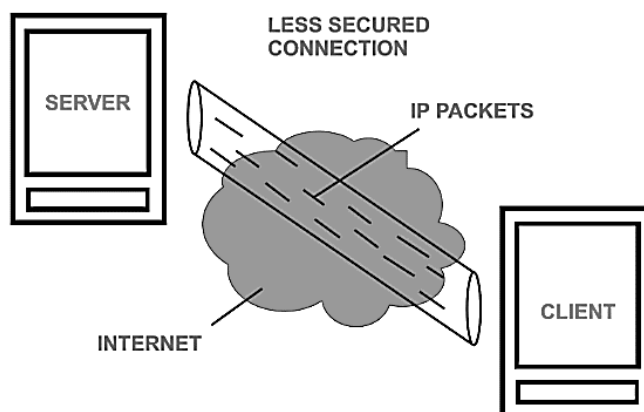


Figure 13. 2 HTTP Connection

History

Tim Berners Lee and his team at CERN gets credit for inventing original HTTP and associated technologies.

HTTP version 0.9 –

This was first version of HTTP which was introduced in 1991.

HTTP version 1.0 –

In 1996, RFC 1945 (Request For Comments) was introduced in HTTP version 1.0.

HTTP version 1.1 –

In January 1997, RFC 2068 was introduced in HTTP version 1.1.

- Improvements and updates to HTTP version 1.1 standard were released under RFC 2616 in June 1999.

HTTP version 2.0 –

The HTTP version 2.0 specification was published as RFC 7540 on May 14, 2015.

HTTP version 3.0 –

HTTP version 3.0 is based on previous RFC draft.

- It is renamed as HyperText Transfer Protocol QUIC which is a transport layer network protocol developed by Google.

History

- The protocols that are used to transfer hypertext between two computers is known as HyperText Transfer Protocol.
- HTTP provides standard between a web browser and web server to establish communication.
- It is set of rules for transferring data from one computer to another.
- Data such as text, images, and other multimedia files are shared on the World Wide Web.
- Whenever a web user opens their web browser, user will indirectly uses HTTP.
- It is an application protocol which is used for distributed, collaborative, hypermedia information systems.

How it works?

- Whenever we want to open any website then first we open web browser after that we will type URL of that website (e.g., www.facebook.com).
- This URL is now sent to Domain Name Server (DNS).
- Then DNS first check records for this URL in their database, then DNS will return IP address to web browser corresponding to this URL.
- Now browser is able to sent request to actual server.
- After server sends data to client, connection will be closed.
- If we want something else from server we should have to re-establish connection between client and server.

Characteristics of HTTP

- HTTP is IP based communication protocol which is used to deliver data from server to client or vice-versa.
- Server processes a request, which is raised by client and also server and client knows each other only during current request and response period.
- Any type of content can be exchanged as long as server and client are compatible with it.
- Once data is exchanged then servers and client are no more connected with each other.

- It is a request and response protocol based on client and server requirements.
- It is connection less protocol because after connection is closed, server does not remember anything about client and client does not remember anything about server.
- It is stateless protocol because both client and server does not expecting anything from each other but they are still able to communicate.

Advantages

- Memory usage and CPU usage are low because of less simultaneous connections.
- Since there are few TCP connections hence network congestion are less.
- Since handshaking is done at initial connection stage, then latency is reduced because there is no further need of handshaking for subsequent requests.
- The error can be reports without closing connection.
- HTTP allows HTTP pipe-lining of request or response.

Disadvantages of HTTP

- HTTP requires high power to establish communication and transfer data.
- HTTP is less secure, because it does not uses any encryption method like https use TLS to encrypt normal http requests and response.
- HTTP is not optimized for cellular phone and it is too gabby.
- HTTP does not offer genuine exchange of data because it is less secure.
- Client does not close connection until it receives complete data from server and hence server needs to wait for data completion and cannot be available for other clients during this time.
- Hypertext Transfer Protocol Secure (HTTPS)
- Hypertext Transfer Protocol Secure (HTTPS) is an extension of the Hypertext Transfer Protocol (HTTP).
- It is used for secure communication.
- In HTTPS, the communication protocol is encrypted using Transport Layer Security.

Table 13. 2 Difference between HTTP and HTTPS

S.No	HTTP	HTTPS
1.	HTTP stands for HyperText Transfer Protocol.	HTTPS for HyperText Transfer Protocol Secure.
2.	In HTTP, URL begins with "http://".	In HTTPS, URL starts with "https://".
3.	HTTP uses port number 80 for communication.	HTTPS uses 443 port number for communication.
4.	HTTP is considered to be unsecure.	HTTPS is considered as secure.
5.	HTTP works at Application Layer.	HTTPS works at Transport Layer.
6.	In HTTP, Encryption is absent.	Encryption is present in HTTPS.
7.	HTTP does not require any certificates.	HTTPS needs SSL Certificates.

HTTP Non-Persistent & Persistent Connection

- The Hypertext Transfer Protocol (HTTP) is an application-level protocol that uses TCP as an underlying transport and typically runs on port 80. HTTP is a stateless protocol i.e. server maintains no information about past client requests.
- Non-persistent and persistent are the two types of HTTP connections used to connect the client with the webserver.
- The non-persistent connection has connection type 1.0, while the persistent connection has connection type 1.1.

Persistent Connection

- A persistent connection takes 2 RTT for the connection and then transfers as many objects, as wanted, over this single connection.
- RTT stands for the round-trip time taken for an object request and then its retrieval.
- In other words, it is the time taken to request the object from the client to the server and then retrieve it from the server back to the client.

Non-Persistent Connection

- The non-persistent connection takes the connection time of $2RTT + \text{file transmission time}$.
- It takes the first RTT (round-trip time) to establish the connection between the server and the client.
- The second RTT is taken to request and return the object. This case stands for a single object transmission.
- After the client receives the object in non-persistent, the connection is immediately closed.
- This is the basic difference between persistent and non-persistent. The persistent connection ensures the transfer of multiple objects over a single connection.

13.5 World Wide Web (WWW)

- The World Wide Web abbreviated as WWW and commonly known as the web.
- The WWW was initiated by CERN (European laboratory for Nuclear Research) in 1989.
- World Wide Web, which is also known as a Web, is a collection of websites or web pages stored in web servers and connected to local computers through the internet.
- These websites contain text pages, digital images, audios, videos, etc.
- Users can access the content of these sites from any part of the world over the internet using their devices such as computers, laptops, cell phones, etc.
- The WWW, along with internet, enables the retrieval and display of text and media to your device.

World Wide Web (WWW) – Historic Facts

- It is a project created, by Timothy Berner's Lee in 1989.
- It was developed for researchers to work together effectively at CERN. is an organization, named World Wide Web Consortium (W3C), was developed for further development in web.
- This organization is directed by Tim Berner's Lee, aka father of web.

World Wide Web – System Architecture:

- The WWW today is a distributed client/server service, in which a client using a browser can access a service using a server. However, the service provided is distributed over many locations called sites.
- From user's point of view, the web consists of a vast, worldwide connection of documents or web pages.
- Each page may contain links to other pages anywhere in the world.

- The pages can be retrieved and viewed by using browsers of which internet explorer, Netscape Navigator, Google, Chrome, etc are the popular ones.
- The browser fetches the page requested interprets the text and formatting commands on it, and displays the page, properly formatted, on the screen.
- The basic model of how the web works is shown in figure below.
- Here the browser is displaying a web page on the client machine.
- When the user clicks on a line of text that is linked to a page on the abd.com server, the browser follows the hyperlink by sending a message to the abd.com server asking it for the page.
- Here the browser displaying web page on the client machine. When the user clicks on a line of text that is linked to a page on abd.com, the Web browser follows the hyperlink by sending a message to abd.com server asking it for the page.

Working of World Wide Web

- The World Wide Web is based on several different technologies : Web browsers, Hypertext Markup Language (HTML) and Hypertext Transfer Protocol (HTTP).
- An Web browser is used to access webpages. Web browsers can be defined as programs which display text, data, pictures, animation and video on the Internet.
- Hyperlinked resources on the World Wide Web can be accessed using software interface provided by Web browsers.
- Initially Web browsers were used only for surfing the Web but now they have become more universal.
- Web browsers can be used for several tasks including conducting searches, mailing, transferring files, and much more.
- Some of the commonly used browsers are Internet Explorer, Opera Mini, Google Chrome.

Features of World Wide Web

- 1) HyperText Information System
- 2) Cross-Platform
- 3) Distributed
- 4) Open Standards and Open Source
- 5) Uses Web Browsers to provide a single interface for many services
- 6) Dynamic, Interactive and Evolving.
- 7) "Web 2.0"

1) HyperText Information System

HyperText is a way to link and access information of various kinds as a web of nodes in which the user can browse at will.

Potentially, HyperText provides a single user-interface to many large classes of stored information, such as reports, notes, data-bases, computer documentation and on-line systems help.

2) Cross-Platform

Developers are increasingly using HTML5, JavaScript and CSS3 to aid in the creation of web apps and native mobile apps.

This process is especially useful when dealing with cross-platform development or when working with content that already exists in some form on the web.

3) Distributed

The web is a distributed system of delivering linked documents over the Internet.

It is called a distributed system because information can reside on different computers around the world. Yet be easily linked together using hypertext.

The web uses hypertext to create links from together using hypertext.

4) Open Standards and Open Source

As one of the important provider of ICT Standards for the past decade, in the area of Web technologies (HTML, URL, XML, http w/IETF, CSS, WAI guidelines, Web Services, Semantic Web, etc),

The international World Wide Web Consortium (W3C) is well positioned to give its opinion on the matter of Open Standards definition.

W3C follows a process that promotes the development of high-quality standards.

This process has evolved over a period of ten years, from a very rough consensus building approach of writing specifications (ancestor IETF model), to a formal set of obligations that promote fairness, responsiveness, and progress: all facets of the W3C mission.

5) Uses Web Browsers to provide a single interface for many services

The World Wide Web lets you access many distinct Internet services through a common set of protocols.

A single application, the web browser (e.g., Mozilla Firefox, Internet Explorer, Lynx), can access most of these services and typically has the ability to launch helper applications or use plug-ins for services and file types it cannot access directly.

6) Dynamic, Interactive and Evolving

The World Wide Web lets you access interactive webpages which provide dynamic web content right from the server databases.

To facilitate it, many client side and server side scripting languages are utilized making it dynamic and interactive.

7) Web2.0

Web 2.0 is the second generation of web. It was defined by Dale Dougherty in 2004 as a read-write web.

The concept began with a conference brainstorming session between O' Reilly and Media live International. The technologies of web 2.0 allow assembling and managing large global crowds with common interests in social interactions.

Tim O'Reilly defines web 2.0 on his website as

"Web 2.0 is the business revolution in the computer industry caused by the move to the internet as platform, and an attempt to understand the rules for success on that new platform".

- Chief among those rules is this. Build applications that harness network effects to get better the more people use them."

Web 2.0 facilitates major properties like participatory, collaborative, and distributed practices which enable formal and in formal spheres of daily activities on going on web.

It resemble major distinct characteristics of Web 2.0 include —relationship" technologies, participatory media and a social digital technology which in term can also defined as the wisdom web.

People-centric web and participative web is taken in to concern and which facilities reading and writing on the web which makes the web transaction bi-directional.

Components of Web

There are 3 components of web:

- 1) Uniform Resource Locator (URL): serves as system for resources on web.
- 2) Hypertext Transfer Protocol (HTTP): specifies communication of browser and server.
- 3) Hyper Text Markup Language (HTML): defines structure, organization and content of webpage.

Virtual Private Network (VPN)

- A virtual private network (VPN) extends a private network across a public network.
- It enables users to send and receive data across shared or public networks as if their computing devices were directly connected to the private network.
- Applications running across a VPN may therefore benefit from the functionality, security, and management of the private network.
- Encryption is a common, although not an inherent, part of a VPN connection.

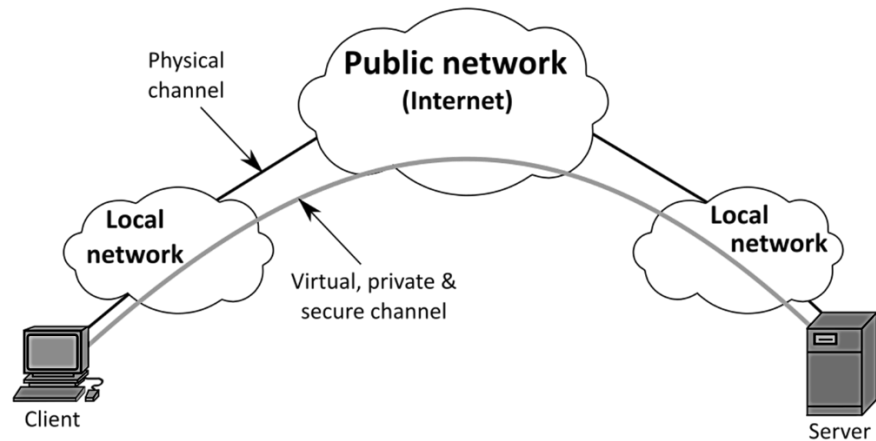


Figure 13.3 Virtual Private Network (VPN)

- VPN technology was developed to provide access to corporate applications and resources to remote users, mobile users, and to branch offices.
- For security, the private network connection may be established using an encrypted layered tunneling protocol, and users may be required to pass various authentication methods to gain access to the VPN.
- In other applications, Internet users may secure their connections with a VPN to circumvent geo-blocking and censorship or to connect to proxy servers to hide their IP address from the target server.
- Some websites, however, block access to known IP addresses used by VPNs to prevent the circumvention of their geo-restrictions, and many VPN providers have been developing strategies to get around these blockades.
- A VPN is created by establishing a virtual point-to-point connection through the use of dedicated circuits or with tunneling protocols over existing networks.
- A VPN available from the public Internet can provide some of the benefits of a wide area network (WAN).
- From a user perspective, the resources available within the private network can be accessed remotely.

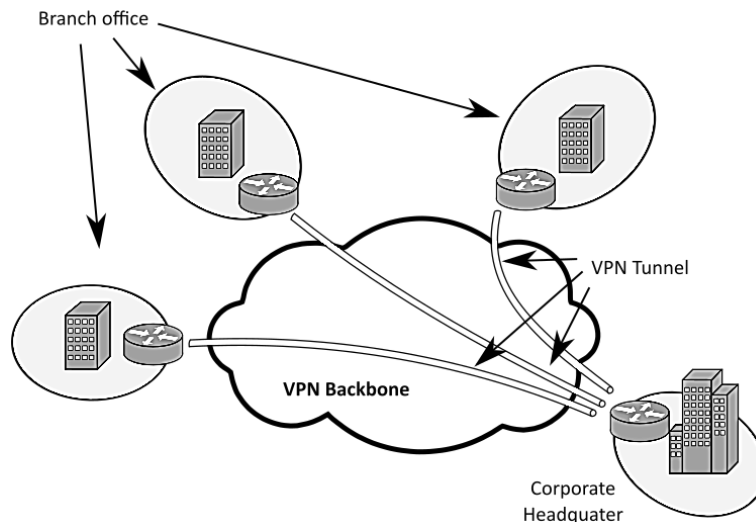


Figure 13.4 Virtual Private Network (VPN)

Types of Virtual Private Network (VPN)

Three broad categories of VPNs exist, namely

- 1) Remote access,
- 2) Intranet-based site-to-site, and
- 3) Extranet-based site-to-site.

While individual users most frequently interact with remote access VPNs, businesses make use of site-to-site VPNs more often.

Types of Virtual Private Network (VPN)

- Early data networks allowed VPN-style connections to remote sites through dial-up modem or through leased line connections utilizing X.25, Frame Relay and

Asynchronous Transfer Mode (ATM) virtual circuits provided through networks owned and operated by telecommunication carriers.

- These networks are not considered true VPNs because they passively secure the data being transmitted by the creation of logical data streams.
- They have been replaced by VPNs based on IP and IP/Multi-protocol Label Switching (MPLS) Networks, due to significant cost-reductions and increased bandwidth provided by new technologies such as digital subscriber line (DSL) and fiber-optic networks.

Virtual Private Network (VPN) Classifications

- VPNs can be characterized as host-to-network or remote access by connecting a single computer to a network or as site-to-site for connecting two networks.
- In a corporate setting, remote-access VPNs allow employees to access the company's intranet from outside the office.
- Site-to-site VPNs allow collaborators in geographically disparate offices to share the same virtual network.
- A VPN can also be used to interconnect two similar networks over a dissimilar intermediate network, such as two IPv6 networks connected over an IPv4 network.

VPN systems may be classified by:

- the tunneling protocol used to tunnel the traffic
- the tunnel's termination point location, e.g., on the customer edge or network-provider edge

- the type of topology of connections, such as site-to-site or network-to-network
- the levels of security provided
- the OSI layer they present to the connecting network, such as Layer 2 circuits or Layer 3 network connectivity
- the number of simultaneous connections

Security Mechanisms

VPNs cannot make online connections completely anonymous, but they can usually increase privacy and security.

To prevent disclosure of private information, VPNs typically allow only authenticated remote access using tunneling protocols and encryption techniques.

The VPN security model provides:

- confidentiality such that even if the network traffic is sniffed at the packet level (see network sniffer and deep packet inspection), an attacker would see only encrypted data
- sender authentication to prevent unauthorized users from accessing the VPN
- message integrity to detect any instances of tampering with transmitted messages.

Life Cycle phases of IPSec Tunnel in VPN

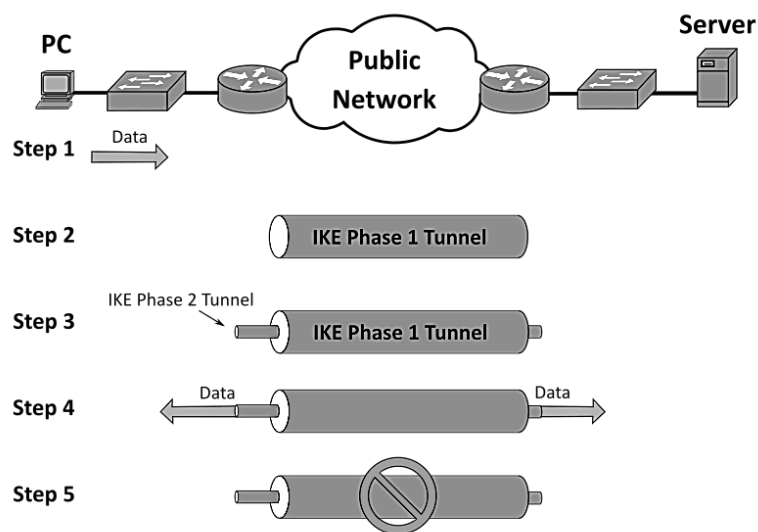


Figure 13.5 Life Cycle phases of IPSec Tunnel in VPN

Secure VPN Protocols

Secure VPN protocols include the following:

- 1) **Internet Protocol Security (IPsec)** was initially developed by the Internet Engineering Task Force (IETF) for IPv6, which was required in all standards-compliant implementations of IPv6 before RFC 6434 made it only a recommendation.
 - This standards-based security protocol is also widely used with IPv4 and the Layer 2 Tunneling Protocol.
 - Its design meets most security goals:

availability,

integrity, and

confidentiality.

- IPsec uses encryption, encapsulating an IP packet inside an IPsec packet.

- De-encapsulation happens at the end of the tunnel, where the original IP packet is decrypted and forwarded to its intended destination.
- 2) **Transport Layer Security (SSL/TLS)** can tunnel an entire network's traffic (as it does in the OpenVPN project and SoftEther VPN project) or secure an individual connection.

A number of vendors provide remote-access VPN capabilities through SSL.

An SSL VPN can connect from locations where IPsec runs into trouble with Network Address Translation and firewall rules.

- 3) **Datagram Transport Layer Security (DTLS)** -
 - used in Cisco AnyConnect VPN and in OpenConnect VPN to solve the issues SSL/TLS has with tunneling over TCP (tunneling TCP over TCP can lead to big delays and connection aborts).
- 4) **Microsoft Point-to-Point Encryption (MPPE)** works with the Point-to-Point Tunneling Protocol and in several compatible implementations on other platforms.
- 5) **Microsoft Secure Socket Tunneling Protocol (SSTP)** tunnels Point-to-Point Protocol (PPP) or Layer 2 Tunneling Protocol traffic through an SSL/TLS channel (SSTP was introduced in Windows Server 2008 and in Windows Vista Service Pack 1).
- 6) **Multi Path Virtual Private Network (MPVPN)**. Ragula Systems Development Company owns the registered trademark "MPVPN".
- 7) **Secure Shell (SSH) VPN** - OpenSSH offers VPN tunneling (distinct from port forwarding) to secure remote connections to a network or to inter-network links. OpenSSH server provides a limited number of concurrent tunnels. The VPN feature itself does not support personal authentication.
- 8) **WireGuard** is a protocol.
 - In 2020, WireGuard support was added to both the Linux and Android kernels, opening it up to adoption by VPN providers.
 - By default, WireGuard utilizes Curve25519 for key exchange and ChaCha20 for encryption, but also includes the ability to pre-share a symmetric key between the client and server.
 - Almost all commercial VPNs adopted this protocol as the default one.

Authentication

- Tunnel endpoints must be authenticated before secure VPN tunnels can be established.
- User-created remote-access VPNs may use passwords, biometrics, two-factor authentication or other cryptographic methods.
- Network-to-network tunnels often use passwords or digital certificates.
- They permanently store the key to allow the tunnel to establish automatically, without intervention from the administrator.

Routing

- Tunneling protocols can operate in a point-to-point network topology that would theoretically not be considered a VPN because a VPN by definition is expected to support arbitrary and changing sets of network nodes.
- But since most router implementations support a software-defined tunnel interface, customer-provisioned VPNs often are simply defined tunnels running conventional routing protocols.

Provider-provisioned VPN Building-Blocks

Tunneling protocols can operate in a point-to-point network topology that would theoretically not be considered a VPN because a VPN by definition is expected to support arbitrary and changing sets of network nodes. But since most router implementations support a software-defined tunnel interface, customer-provisioned VPNs often are simply defined tunnels running conventional routing protocols.

Depending on whether a provider-provisioned VPN (PPVPN) operates in layer 2 or layer 3, the building blocks described below may be L2 only, L3 only, or a combination of both. Multi-protocol label switching (MPLS) functionality blurs the L2-L3 identity.

RFC 4026 generalized the following terms to cover L2 MPLS VPNs and L3 (BGP) VPNs, but they were introduced in RFC 2547.

Customer (C) Devices

A device that is within a customer's network and not directly connected to the service provider's network. C devices are not aware of the VPN.

Customer Edge Device (CE)

A device at the edge of the customer's network which provides access to the PPVPN. Sometimes it is just a demarcation point between provider and customer responsibility. Other providers allow customers to configure it.

Provider Edge Device (PE)

A device, or set of devices, at the edge of the provider network which connects to customer networks through CE devices and presents the provider's view of the customer site. PEs are aware of the VPNs that connect through them, and maintain VPN state.

Provider Device (P)

A device that operates inside the provider's core network and does not directly interface to any customer endpoint. It might, for example, provide routing for many provider-operated tunnels that belong to different customers' PPVPNs. While the P device is a key part of implementing PPVPNs, it is not itself VPN-aware and does not maintain VPN state. Its principal role is allowing the service provider to scale its PPVPN offerings, for example, by acting as an aggregation point for multiple PEs. P-to-P connections, in such a role, often are high-capacity optical links between major locations of providers.

IPsec

The IP security (IPSec) is an Internet Engineering Task Force (IETF) standard suite of protocols between 2 communication points across the IP network that provide data authentication, integrity, and confidentiality. It also defines the encrypted, decrypted and authenticated packets. The protocols needed for secure key exchange and key management are defined in it.

Uses of IPsec

IPsec can be used to do the following things:

- To encrypt application layer data.
- To provide security for routers sending routing data across the public internet.
- To provide authentication without encryption, like to authenticate that the data originates from a known sender.
- To protect network data by setting up circuits using IPsec tunneling in which all data is being sent between the two endpoints is encrypted, as with a Virtual Private Network (VPN) connection.

Components of IP Security

It has the following components:

- 1) Encapsulating Security Payload (ESP)
- 2) Authentication Header (AH)
- 3) Internet Key Exchange (IKE) –

1) Encapsulating Security Payload (ESP) –

It provides data integrity, encryption, authentication and anti replay. It also provides authentication for payload.

2) Authentication Header (AH) –

It also provides data integrity, authentication and anti replay and it does not provide encryption. The anti replay protection, protects against unauthorized transmission of packets. It does not protect data's confidentiality.

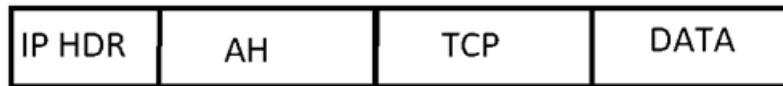


Figure 13. 6 Components of IP Security

3) Internet Key Exchange (IKE) -

- It is a network security protocol designed to dynamically exchange encryption keys and find a way over Security Association (SA) between 2 devices. The Security Association (SA) establishes shared security attributes between 2 network entities to support secure communication.
- The Key Management Protocol (ISAKMP) and Internet Security Association which provides a framework for authentication and key exchange. ISAKMP tells how the set up of the Security Associations (SAs) and how direct connections between two hosts that are using IPsec.
- Internet Key Exchange (IKE) provides message content protection and also an open frame for implementing standard algorithms such as SHA and MD5. The algorithm's IP sec users produces a unique identifier for each packet. This identifier then allows a device to determine whether a packet has been correct or not. Packets which are not authorized are discarded and not given to receiver.

Components of IP Security

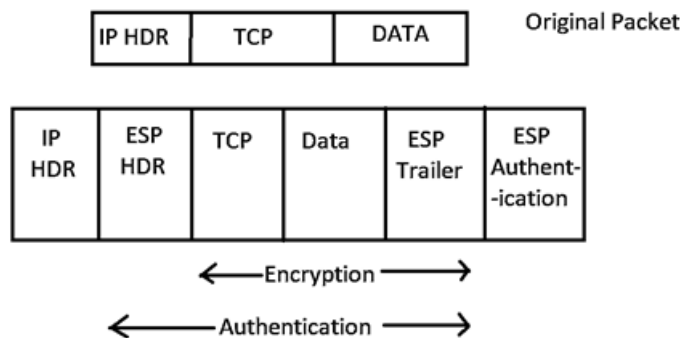


Figure 13. 7 Components of IP Security with Header and Trailer

Working of IP Security

- 1) The host checks if the packet should be transmitted using IPsec or not.

These packet traffic triggers the security policy for themselves. This is done when the system sending the packet apply an appropriate encryption.

The incoming packets are also checked by the host that they are encrypted properly or not.

- 2) Then the IKE Phase 1 starts in which the 2 hosts(using IPsec) authenticate themselves to each other to start a secure channel.

It has 2 modes.

The Main mode which provides the greater security and the Aggressive mode which enables the host to establish an IPsec circuit more quickly.

- 3) The channel created in the last step is then used to securely negotiate the way the IP circuit will encrypt data across the IP circuit.

- 4) Now, the IKE Phase 2 is conducted over the secure channel in which the two hosts negotiate the type of cryptographic algorithms to use on the session and agreeing on secret keying material to be used with those algorithms.
- 5) Then the data is exchanged across the newly created IPsec encrypted tunnel. These packets are encrypted and decrypted by the hosts using IPsec SAs.
- 6) When the communication between the hosts is completed or the session times out then the IPsec tunnel is terminated by discarding the keys by both the hosts.

Summary

Congestion is any stage, into which we aren't able to carry on the normal communication. As you can see on the screen. This is a typical example of a traffic congestion, where there is converging traffic. And at any particular point of time there is a choking situation in the traffic. So the traffic is not having a smooth flow and this is what we mean by congestion. So, it is the reduced quality of service that occurs when a network node or link is carrying more data than it can handle. So when will a traffic overflow occur. A traffic overflow would only occur when a particular road is carrying traffic more than its own capacity. If it is carrying traffic, more than the capacity. It is only under such circumstances that an overflow or a congestion situation can occur. Now typical effects of this congestion include queuing delay. That means you will have long waiting in the queue packet loss packets might be lost during transmission or blocking of new connections, so new connections will be blocked, they will not be able to traverse on that particular network. So, this is about congestion. So, what is a consequence of congestion, well, a consequence of congestion, is that an incremental increase in offered load, leads, either only to a small increase, or even a decrease in the network throughput. So, the performance of the network depreciates. Now, quality of service requirements can be classified or specified as belonging to the four factors. So, quality of service can be specified as delay. Delay variation which I've already told you it is called jitter. It could be on the basis of throughput, that that is the output that I get after a particular time period, and the error rate which is there, and there are two types of quality of service solutions, which I, which are available for us. One is the stateless solution. Another is the stateful solution. So. what are these, let us see one by one. In stateless solution routers maintain no fine-grained state about traffic. So one positive factor of it is that it is scalable, and it is robust, but it has weak services, as there is no guarantee about the kind of delay or performance in a particular application, which we have to encounter, talking about the integrated services, or the in itself, what exactly do we mean by that. Let us now see an architecture for providing quality of service guarantees in IP networks for individual application sessions. So, the architecture, which is to provide quality of service guarantee, and in, in the IP network because it is the Internet Protocol networks that we use so individual application sessions need to take care of this. It relies on the resource reservation and routers need to maintain state information of allocated resources.

Keywords

Stateless Solutions - In stateless solutions Routers maintain no fine grained state about traffic, one positive factor of it is that it is scalable and robust. But it has weak services as there is no guarantee about kind of delay or performance in a particular application which we have to encounter.

Stateful Solutions - In stateful solutions Routers maintain per flow state as flow is very important in providing the Quality-of-Service. i.e. providing powerful services such as guaranteed services and high resource utilization, provides protection and is much less scalable and robust..

A constant-bit-rate (CBR), or a fixed-rate, traffic model has a data rate that does not change. In this type of flow, the average data rate and the peak data rate are the same. Local Area Network: A LAN is a form of local (limited distance), shared packet network for computer communications.

Leaky Bucket Algorithm: It is a network with variable input but output at a constant data rate

Self-Assessment

Multiple Choice Questions

- 1) Which of the following is not true regarding Internet?
 - a) Internet uses TCP/IP protocol suite to connect computer users worldwide.
 - b) Internet was first connected in 1969
 - c) It was developed by ICANN (Internet Corporation for Assigned Names and Numbers) located in the USA.
 - d) The internet works with the help of clients and servers
- 2) When you turn on your computer and type a domain name in the browser search bar, _____?
 - a) It creates a virtual connection between the server and the computer.
 - b) Your browser sends a request to the HTTP server and opens the webpage for you.
 - c) The request is forwarded to the Internet Service Provider who opens the requested page on the computer.
 - d) Your browser sends a request to the DNS server to get the corresponding IP address, and then forwards the request to the respective server.
- 3) Which of the following is not related to the Web?
 - a) It is a model for sharing information
 - b) It is accessed by the Web Browser.
 - c) It makes use of the HTTP protocol
 - d) It is also known as Network of Networks.
- 4) Which of the following is not true regarding URI?
 - a) URI stands for 'Uniform Resource Identifier'.
 - b) A URI can be a name, locator, or both for an online resource.
 - c) URIs are a subset of URLs.
 - d) A URL (a subset of URI) is human-readable text that was designed to replace the IP addresses that computers use to communicate with servers.
- 5) What is not true regarding HTTP?
 - a) HTTP stands for Hyper Text Transfer Protocol.
 - b) It is invented by Charles Babbage.
 - c) Hyper-Text is the type of text which is specially coded with the help of some standard coding language called as HyperText Markup Language (HTML).
 - d) HTTP is set of rules for transferring data from one computer to another.
- 6) Which of the following statements is not True?
 - a) Any type of content can be exchanged with the help of HTTP, as long as server and client are compatible with it.
 - b) Once data is exchanged then servers and client still stay connected with each other.
 - c) It is a request and response protocol based on client and server requirements.
 - d) None of the given choices.
- 7) HTTP is connection less protocol because after connection is closed _____
 - a) Server does not remember anything about client and client does not remember anything about server.

- b) Server always remembers everything about client and client remembers everything about server.
 - c) Server does not remember anything about client and the client remembers everything about server.
 - d) Server always remembers everything about client and client does not remember anything about server.
- 8) Which of the following statements is not true regarding HTTP Non-Persistent connection?
- a) The non-persistent connection has connection type 1.0
 - b) It ensures the transfer of multiple objects over a single connection.
 - c) It takes the connection time of $2RTT + \text{file transmission time}$.
 - d) It takes the first RTT (round-trip time) to establish the connection between the server and the client
- 9) Which of the following is not True?
- a) Web is a collection of websites or web pages stored in web servers and connected to local computers through the internet.
 - b) WWW, along with internet, enables the retrieval and display of text and media to your device.
 - c) WWW is a project created, by Charles Babbage in 1989.
 - d) An organization, named World Wide Web Consortium (W3C), was developed for development in web.
- 10) The WWW today is a client-server service, in which a client using a browser can access a service using a server.
- a) limited
 - b) distributed
 - c) vast
 - d) None of the given choices
- 11) In a URL, the ____ is the computer on which the information is located.
- a) path
 - b) protocol
 - c) host
 - d) None of the given choices
- 12) In a URL, the _____ is the full name of the file where the information is located.
- a) path
 - b) protocol
 - c) host
 - d) None of the given choices
- 13) Which of the following is not a characteristic of a VPN?
- a) It is a secure network
 - b) It is deployed over a shared infrastructure
 - c) It may use tunneling techniques
 - d) It does not provide any cost savings to alternate connectivity options

- 14) What would be a good characterization of a VPN tunnel established between a telecommuter's PC using a VPN client software and a VPN Concentrator at the HQ location?
- a) Remote access VPN
 - b) Site to site VPN
 - c) Extranet VPN
 - d) LAN to LAN VPN
- 15) Which of the following security technique provides confidentiality (data privacy) service?
- a) Hashing
 - b) Key exchange
 - c) Encryption
 - d) All the given choices

Answers: Self-Assessment

- | | | | | |
|-------|-------|-------|-------|-------|
| 1. C | 2. D | 3. D | 4. C | 5. B |
| 6. B | 7. A | 8. B | 9. C | 10. B |
| 11. C | 12. A | 13. D | 14. A | 15. C |

Review Questions

1. Explain the general principles of congestion.
2. What do you understand by QoS? Describe the basic QoS structure.
3. Discuss the following two algorithms:
 - (a) Leaky Bucket
 - (b) Token Bucket
4. What are two types of congestion control? Where is congestion control implemented in each case?
5. Discuss the various causes of the costs of Congestion.

Further Readings



Andrew S. Tanenbaum, *Computer Networks*, Prentice Hall

Behrouz A. Forouzan and Sophia Chung Fegan, *Data Communications and Networking*, McGraw-Hill Companies

Burton, Bill, *Remote Access for Cisco Networks*, McGraw-Hill Osborne Media McGraw-Hill Osborne Media

Rajneesh Agrawal and Bharat Bhushan Tiwari, *Computer Networks and Internet*, Vikas Publication

William A Shay, *Understanding Communication and Networks*, 3rd Edition, Thomson Press.



<https://www.geeksforgeeks.org/basics-computer-networking/>

Unit 14: Network Security

CONTENTS

Objectives

Introduction

14.1 Network Security

14.2 Cyber Security

14.3 Cyber Security Goals

14.4 Types of Cyber Attacks

14.5 Cryptography

14.6 Email

14.7 Firewalls

Summary

Self Assessment

Answer for Self Assessment

Review Questions

Further Readings

Objectives

After this lecture, you would be able to:

- understand the objectives and goals of cybersecurity
- learn about cyber-attacks and its various classifications
- understand the basic principles and types of cryptography
- learn the various ways to ensure message integrity
- understand the cryptographic hash functions for ensuring message integrity
- learn the key aspects related to email security
- learn the various steps involved in implementing the privacy enhanced mail.
- understand the various benefits of using firewalls
- learn about the various types of Firewalls

Introduction

Over the past several years, the world has become interconnected in ways not previously imaginable. Small and large companies have presence on WWW and their offices spread across the globes have inter-office collaboration on a daily basis. Hence, all of these interconnections rely in large part on our ability to protect the networks that create those connections. Network security is a broad topic with multi-layered approach. It can be addressed at the data link layer, network layer and application layer. The issues concerned are: packet intrusion and encryption, IP packets and routing tables with their update version, and host-level bugs occurred at data link layer, network layer and application respectively.

The TCP/IP protocols are being used globally irrespective of the nature of the organizations whether it belongs to general category of organizations or security specific sensitive organizations. The news or information about hacking of some web site or portal by some undesired people is

very common nowadays. This shows that TCP/IP protocols are susceptible to intercept. This generated a need to ensure all round security for the network in an organization. The task of network administrator had to widen to include the overall security of the network. He must ensure that all parts of this network are adequately protected, and adequate measures of security have been implemented within a TCP/IP network. He should be aware of an effective security policy. He should also be able to pinpoint the main areas of risk that the network may face. Basically, these main areas of risk vary from network to network depending upon the organization functioning. There are therefore various security related aspects, which have direct implications for network administrator along with the means to monitor the implemented measures of security effectively and to tackle the problem of breach of security if it happens.

14.1 Network Security

The main objective of the network is to share information among its users situated locally or remotely. Therefore, it is possible that undesired user can hack the network and can prove to be harmful for the health of the network or user. There are few basic points, which must be followed by network administrator to provide the network an adequate security other than network specific security as in case of e-commerce, etc. These are given below:

- Networks are designed to share information. Therefore, the network must be clearly configured to identify the shareable information and non-shareable information.
- The network should also clear with whom the shareable information could be shared.
- With the increase of system security, the price for its management will also increase; accordingly, therefore a compromising level between security and prices should be established as per the requirement of the network security system policy. This will largely depend upon the level of security needed to apply in the network, overall security requirements and the effective implementation of chosen level of security.
- Division of the responsibilities concerning the network security must be clearly defined between users and system administrator.
- The requirements for security must be detailed within a network security policy of the organization that indicates the valuable data and their associated cost to the business.
- After defining the detailed network security policy and identifying the clear cut responsibilities in the organization, the system administrator should be made then responsible for ensuring that the security policy is effectively applied to the company environment, including the existing networking infrastructure.

14.2 Cyber Security

"Cybersecurity is primarily about people, processes, and technologies working together to encompass the full range of threat reduction, vulnerability reduction, deterrence, international engagement, incident response, resiliency, and recovery policies and activities, including computer network operations, information assurance, law enforcement, etc.". Cybersecurity is the protection of Internet-connected systems, including hardware, software, and data from cyber-attacks. It is made up of two words one is cyber and other is security. Cyber is related to the technology which contains systems, network and programs or data. Whereas security related to the protection which includes systems security, network security and application and information security. It is the body of technologies, processes, and practices designed to protect networks, devices, programs, and data from attack, theft, damage, modification or unauthorized access. It may also be referred to as information technology security. We can also define cybersecurity as the set of principles and practices designed to protect our computing resources and online information against threats. Due to the heavy dependency on computers in a modern industry that store and transmit an abundance of confidential and essential information about the people, cybersecurity is a critical function and needed insurance of many businesses.

14.3 Cyber Security Goals

The objective of Cybersecurity is to protect information from being stolen, compromised or attacked.

Cybersecurity can be measured by at least one of three goals-

- Protect the confidentiality of data.
- Preserve the integrity of data.
- Promote the availability of data for authorized users.

These goals form the confidentiality, integrity, availability (CIA) triad, the basis of all security programs. The CIA triad is a security model that is designed to guide policies for information security within the premises of an organization or company. This model is also referred to as the AIC (Availability, Integrity, and Confidentiality) triad to avoid the confusion with the Central Intelligence Agency. The elements of the triad are considered the three most crucial components of security. The CIA criteria are one that most of the organizations and companies use when they have installed a new application, creates a database or when guaranteeing access to some data. For data to be completely secure, all of these security goals must come into effect. These are security policies that all work together, and therefore it can be wrong to overlook one policy.

Cyber Security Goals

1. **Confidentiality** - is roughly equivalent to privacy and avoids the unauthorized disclosure of information. It involves the protection of data, providing access for those who are allowed to see it while disallowing others from learning anything about its content. It prevents essential information from reaching the wrong people while making sure that the right people can get it. Data encryption is a good example to ensure confidentiality.

Tools for Confidentiality

- a) **Encryption** - is a method of transforming information to make it unreadable for unauthorized users by using an algorithm. The transformation of data uses a secret key (an encryption key) so that the transformed data can only be read by using another secret key (decryption key). It protects sensitive data such as credit card numbers by encoding and transforming data into unreadable cipher text. This encrypted data can only be read by decrypting it. Asymmetric-key and symmetric-key are the two primary types of encryption.
- b) **Access control** - Access control defines rules and policies for limiting access to a system or to physical or virtual resources. It is a process by which users are granted access and certain privileges to systems, resources or information.
- c) **Authentication** - A process that ensures and confirms a user's identity or role that someone has. Authentication is the necessity of every organizations because it enables organizations to keep their networks secure by permitting only authenticated users to access its protected resources. These resources may include computer systems, networks, databases, websites and other network-based applications or services. It can be done in a number of different ways, but it is usually based on a combination of-
 - something the person has (like a smart card or a radio key for storing secret keys),
 - something the person knows (like a password),
 - something the person is (like a human with a fingerprint).
- d) **Authorization** - is a security mechanism which gives permission to do or have something. It is **used** to determine a person or system is allowed access to resources, based on an access control policy, including computer programs, files, services, data and application features. It is normally preceded by authentication for user identity verification. System administrators are typically assigned permission levels covering all system and user resources. During authorization, a system verifies an authenticated user's access rules and either grants or refuses resource access.
- e) **Physical Security** - describes measures designed to deny the unauthorized access of IT assets like facilities, equipment, personnel, resources and other properties from damage. It protects these assets from physical threats including theft, vandalism, fire and natural disasters.

2. Integrity

Integrity refers to the methods for ensuring that data is real, accurate and safeguarded from unauthorized user modification. It is the property that information has not be altered in an unauthorized way, and that source of the information is genuine.

Tools for Integrity

- a) **Backups** - is the periodic archiving of data. It is a process of making copies of data or data files to use in the event when the original data or data files are lost or destroyed. It is also used to make copies for historical purposes, such as for longitudinal studies, statistics or for

historical records or to meet the requirements of a data retention policy. Many applications especially in a Windows environment, produce backup files using the .BAK file extension.

- b) **Checksums** - is a numerical value used to verify the integrity of a file or a data transfer. In other words, it is the computation of a function that maps the contents of a file to a numerical value. They are typically used to compare two sets of data to make sure that they are the same. A checksum function depends on the entire contents of a file. It is designed in a way that even a small change to the input file (such as flipping a single bit) likely to results in different output value.
- c) **Data Correcting Codes** - It is a method for storing data in such a way that small changes can be easily detected and automatically corrected.
- d) **Availability** - Availability is the property in which information is accessible and modifiable in a timely fashion by those authorized to do so. It is the guarantee of reliable and constant access to our sensitive data by authorized people.

Tools for Availability

- 1) **Physical Protections** - Physical safeguard means to keep information available even in the event of physical challenges. It ensure sensitive information and critical information technology are housed in secure areas.
- 2) **Computational Redundancies** - It is applied as fault tolerant against accidental faults. It protects computers and storage devices that serve as fallbacks in the case of failures.

14.4 Types of Cyber Attacks

A cyber-attack is an exploitation of computer systems and networks. It uses malicious code to alter computer code, logic or data and lead to cybercrimes, such as information and identity theft. We are living in a digital era. Now a day, most of the people use computer and internet. Due to the dependency on digital things, the illegal computer activity is growing and changing like any type of crime.

Classifications of Cyber Attacks

1. Web-based Attacks

These are the attacks which occur on a website or web applications. Some of the important web-based attacks are as follows-

- a) **Injection attacks** - It is the attack in which some data will be injected into a web application to manipulate the application and fetch the required information.
- b) **DNS Spoofing** - DNS Spoofing is a type of computer security hacking. Whereby a data is introduced into a DNS resolver's cache causing the name server to return an incorrect IP address, diverting traffic to the attacker's computer or any other computer.
- c) **Session Hijacking** - It is a security attack on a user session over a protected network. Web applications create cookies to store the state and user sessions. By stealing the cookies, an attacker can have access to all of the user data.
- d) **Phishing** - Phishing is a type of attack which attempts to steal sensitive information like user login credentials and credit card number. It occurs when an attacker is masquerading as a trustworthy entity in electronic communication.
- e) **Brute force** - It is a type of attack which uses a trial and error method. This attack generates a large number of guesses and validates them to obtain actual data like user password and personal identification number. This attack may be used by criminals to crack encrypted data, or by security, analysts to test an organization's network security.
- f) **Denial of Service** - It is an attack which meant to make a server or network resource unavailable to the users. It accomplishes this by flooding the target with traffic or sending it

information that triggers a crash. It uses the single system and single internet connection to attack a server.

Denial of Service can be classified into the following-

- **Volume-based attacks-** Its goal is to saturate the bandwidth of the attacked site and is measured in bit per second.
- **Protocol attacks-** It consumes actual server resources and is measured in a packet.
- **Application layer attacks-** Its goal is to crash the web server and is measured in request per second.

- g) **Dictionary attacks** - This type of attack stored the list of a commonly used password and validated them to get original password.
- h) **URL Interpretation** - It is a type of attack where we can change the certain parts of a URL, and one can make a web server to deliver web pages for which he is not authorized to browse.
- i) **File Inclusion attacks-** It is a type of attack that allows an attacker to access unauthorized or essential files which is available on the web server or to execute malicious files on the web server by making use of the include functionality.
- j) **Man in the middle attacks** - It is a type of attack that allows an attacker to intercept the connection between client and server and acts as a bridge between them. Due to this, an attacker will be able to read, insert and modify the data in the intercepted connection.

2. **System-Based Attacks** - These are the attacks which are intended to compromise a computer or a computer network. Some of the important system-based attacks are as follows-

- a) **Virus** - It is a type of malicious software program that spread throughout the computer files without the knowledge of a user. It is a self-replicating malicious computer program that replicates by inserting copies of itself into other computer programs when executed.
- b) **Worm** - It is a type of malware whose primary function is to replicate itself to spread to uninfected computers. It works same as the computer virus. Worms often originate from email attachments that appear to be from trusted senders.
- c) **Trojan horse** - It is a malicious program that occurs unexpected changes to computer setting and unusual activity, even when the computer should be idle. It misleads the user of its true intent. It appears to be a normal application but when opened/executed some malicious code will run in the background.
- d) **Backdoors** - It is a method that bypasses the normal authentication process. A developer may create a backdoor so that an application or operating system can be accessed for troubleshooting or other purposes.
- e) **Bots** - A bot (short for "robot") is an automated process that interacts with other network services. Some bots program run automatically, while others only execute commands when they receive specific input. Common examples of bots program are the crawler, chatroom bots, and malicious bots.

14.5 Cryptography

Cryptography is the study and practice of techniques for secure communication in the presence of third parties called adversaries. It deals with developing and analyzing protocols which prevents malicious third parties from retrieving information being shared between two entities thereby following the various aspects of information security. Secure Communication refers to the scenario where the message or data shared between two parties can't be accessed by an adversary. In

Cryptography, an Adversary is a malicious entity, which aims to retrieve precious information or data thereby undermining the principles of information security.

The core principles of modern-day cryptography are:

- 1) Data Confidentiality refers to certain rules and guidelines usually executed under confidentiality agreements which ensure that the information is restricted to certain people or places.
- 2) Data integrity refers to maintaining and making sure that the data stays accurate and consistent over its entire life cycle.
- 3) Authentication is the process of making sure that the piece of data being claimed by the user belongs to it.
- 4) Non-repudiation refers to ability to make sure that a person or a party associated with a contract or a communication cannot deny the authenticity of their signature over their document or the sending of a message.

Consider two parties Alice and Bob. Now, Alice wants to send a message m to Bob over a secure channel. The sender's message or sometimes called the Plaintext, is converted into an unreadable form using a Key k . The resultant text obtained is called the Ciphertext. This process is known as Encryption. At the time of receiving, the Ciphertext is converted back into the plaintext using the same Key k , so that it can be read by the receiver. This process is known as Decryption.

Alice (Sender) Bob (Receiver)

$$C = E(m, k) \longrightarrow m = D(C, k)$$

Here, C refers to the Ciphertext while E and D are the Encryption and Decryption algorithms respectively. Let's consider the case of Caesar Cipher or Shift Cipher as an example. As the name suggests, in Caesar Cipher each character in a word is replaced by another character under some defined rules. Thus, if A is replaced by D , B by E and so on. Then, each character in the word would be shifted by a position of 3.



Plaintext : ABCDE

Ciphertext : DEFGH



Even if the adversary knows that the cipher is based on Caesar Cipher, it cannot predict the plaintext as it doesn't have the key in this case which is to shift the characters back by three places.

Cryptography Basic Principles

Whenever we come across the term cryptography, the first thing and probably the only thing that comes to our mind is private communication through encryption. There is more to cryptography than just encryption. In this article, we will try to learn the basics of cryptography.

The Basic Principles are:

- 1) **Encryption** - In a simplest form, encryption is to convert the data in some unreadable form. This helps in protecting the privacy while sending the data from sender to receiver. On the receiver side, the data can be decrypted and can be brought back to its original form. The reverse of encryption is called as decryption. The concept of encryption and decryption requires some extra information for encrypting and decrypting the data. This information is known as key. There may be cases when same key can be used for both encryption and decryption while in certain cases, encryption and decryption may require different keys.
- 2) **Authentication** - This is another important principle of cryptography. In a layman's term, authentication ensures that the message was originated from the originator claimed in the message. Now, one may think how to make it possible? Suppose, Alice sends a message to Bob and now Bob wants proof that the message has been indeed sent by Alice. This can be made possible if Alice performs some action on message that Bob knows only Alice can do. Well, this forms the basic fundamental of Authentication.
- 3) **Integrity** - Now, one problem that a communication system can face is the loss of integrity of messages being sent from sender to receiver. This means that Cryptography should ensure that

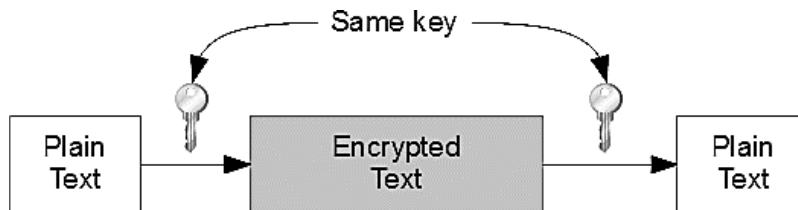
the messages that are received by the receiver are not altered anywhere on the communication path. This can be achieved by using the concept of cryptographic hash.

- 4) **Non Repudiation** - What happens if Alice sends a message to Bob but denies that she has actually sent the message? Cases like these may happen and cryptography should prevent the originator or sender to act this way. One popular way to achieve this is through the use of digital signatures.

Types of Cryptography

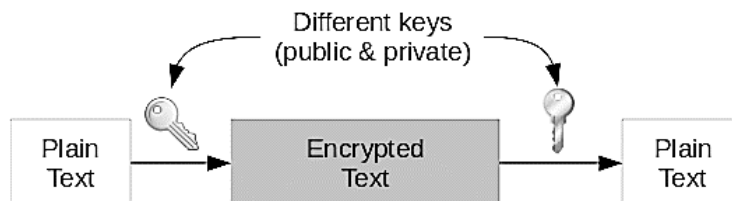
There are three types of cryptography techniques:

- 1) Secret key Cryptography
 - 2) Public key cryptography
 - 3) Hash Functions
- 1) **Secret Key Cryptography** - This type of cryptography technique uses just a single key. The sender applies a key to encrypt a message while the receiver applies the same key to decrypt the message. Since only single key is used so we say that this is a symmetric encryption.



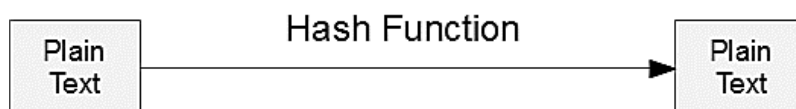
The biggest problem with this technique is the distribution of key as this algorithm makes use of single key for encryption or decryption.

- 2) **Public Key Cryptography** - This type of cryptography technique involves two key crypto system in which a secure communication can take place between receiver and sender over insecure communication channel. Since a pair of keys is applied here so this technique is also known as asymmetric encryption.



In this method, each party has a private key and a public key. The private is secret and is not revealed while the public key is shared with all those whom you want to communicate with. If Alice wants to send a message to bob, then Alice will encrypt it with Bob's public key and Bob can decrypt the message with its private key. This is what we use when we setup public key authentication in opens to login from one server to another server in the backend without having to enter the password.

- 3) **Hash Functions** - This technique does not involve any key. Rather it uses a fixed length hash value that is computed on the basis of the plain text message. Hash functions are used to check the integrity of the message to ensure that the message has not be altered, compromised or affected by virus.



So we see that how different types of cryptography techniques are used to implement the basic principles that we discussed earlier.

Cryptography

Let's consider the case of Caesar Cipher or Shift Cipher as an example. As the name suggests, in Caesar Cipher each character in a word is replaced by another character under some defined rules. Thus, if A is replaced by D, B by E and so on. Then, each character in the word would be shifted by a position of 3.



Plaintext : ABCDE

Ciphertext : DEFGH



Note that even if the adversary knows that the cipher is based on Caesar Cipher, it cannot predict the plaintext as it doesn't have the key in this case which is to shift the characters back by three places.

Message Integrity

The cryptography systems that we have studied so far provide secrecy, or confidentiality, but not integrity. However, there are occasions where we may not even need secrecy but instead must have integrity. Message Integrity describes the concept of ensuring that data has not been modified/tampered with/alterd in transit.



If Bob receives a message (either be encrypted or be in plaintext) from Alice, he needs to verify:

- 1) The message indeed originated from Alice.
- 2) The message was not tempered with on its way to Bob.

Ways to Ensure Message Integrity

1. Documents & Fingerprints
2. Message and Message Digest

1. Documents & Fingerprints

One way to preserve the integrity of a document is using a fingerprint. If Alice needs to be sure that the contents of her document will not be changed, she can put her fingerprint at the bottom of the document.

2. Message & Message Digest

The electronic equivalent of the document and fingerprint pair is the message and digest pair.

Figure: Message and Digest

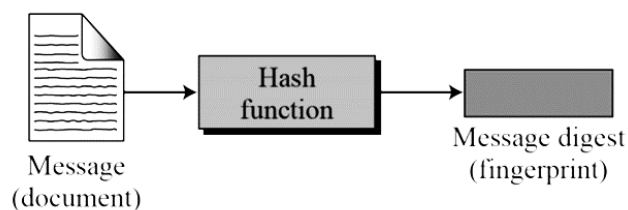


Figure: Message Digest

Differences between Methods of Message Integrity

The two pairs (document/fingerprint) and (message/message digest) have some differences.

- The document and fingerprint are physically linked together.
- The message and message digest can be unlinked separately, and, most importantly, the message digest needs to be safe from change.
- It is important to note that the message digest needs to be safe from change.

Checking Integrity

It is important to note that the message digest needs to be safe from change.

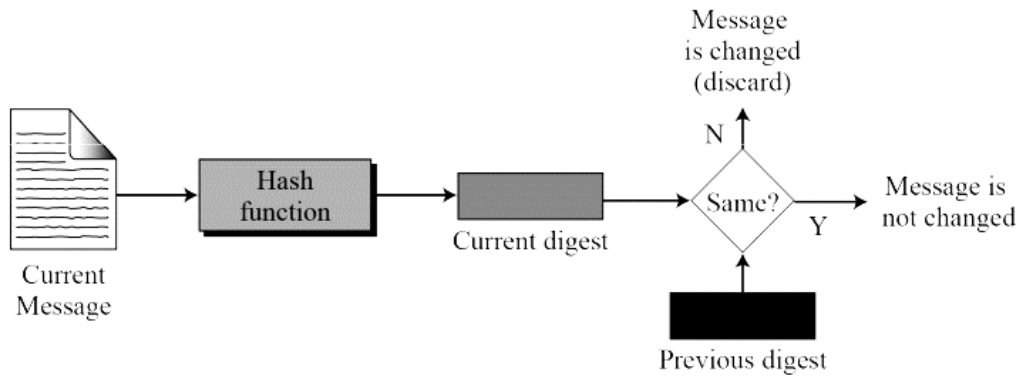


Figure: Message Integrity

Cryptographic Hash Function Criteria

It is important to note that the message digest needs to be safe from change. A cryptographic hash function must satisfy three criteria:

- 1) Preimage Resistance.
- 2) Second Preimage Resistance.
- 3) Collision Resistance.

- 1) **Preimage Resistance:** The hash function must be a one-way function:

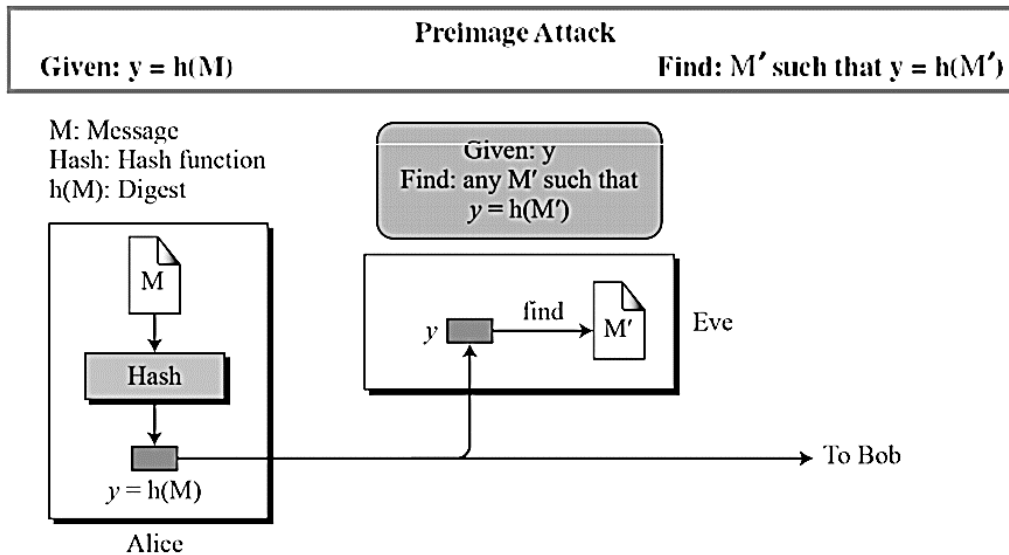


Figure: First preimage

For any given code h , it is computationally infeasible to find h^{-1} .

- 2) **Second Preimage Resistance:**

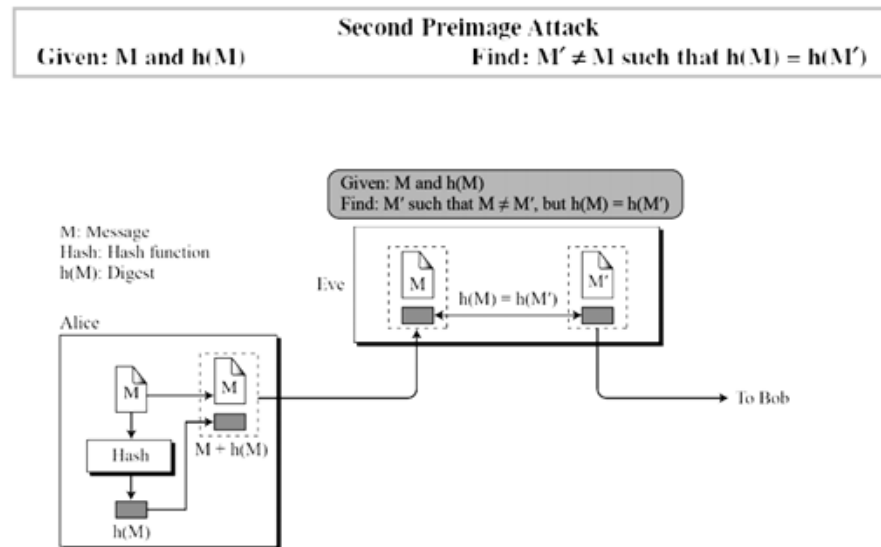


Figure: Second preimage

3) Collision Resistance

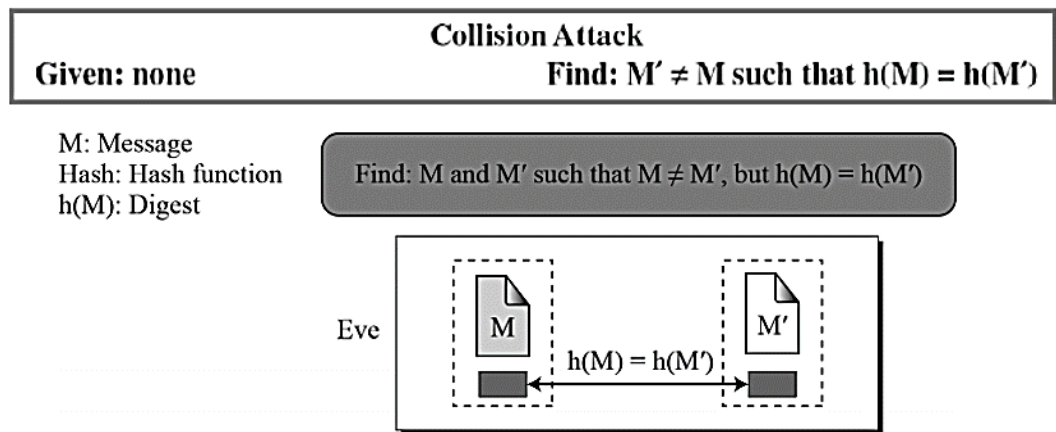


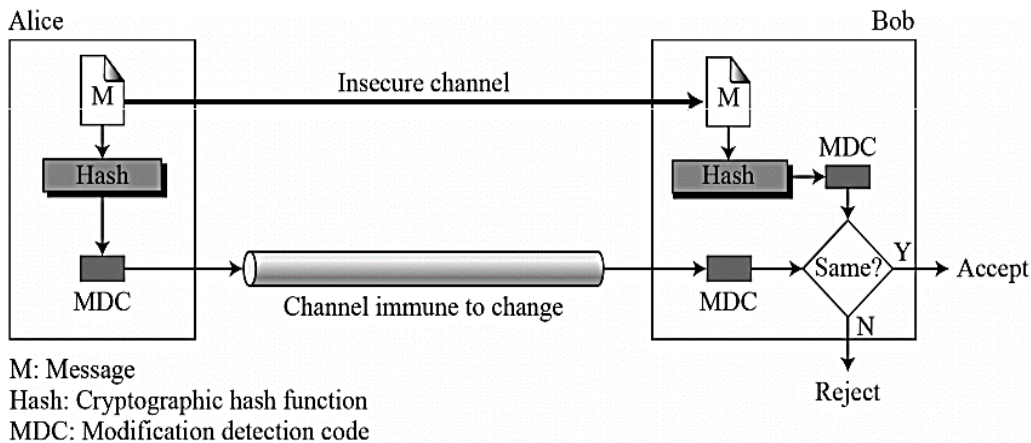
Figure : Collision

Message Authentication

A message digest does not authenticate the sender of the message. To provide message authentication, Alice needs to provide proof that it is Alice sending the message and not an impostor. The digest created by a cryptographic hash function is normally called a modification detection code (MDC). What we need for message authentication is a message authentication code (MAC).

Modification Detection Code (MDC)

A modification detection code (MDC) is a message digest that can prove the integrity of the message: that message has not been changed. If Alice needs to send a message to Bob and be sure that the message will not change during transmission, Alice can create a message digest, MDC, and send both the message and the MDC to Bob. Bob can create a new MDC from the message and compare the received MDC and the new MDC. If they are the same, the message has not been changed.



The difference between MDC and MAC is that the second include A secrete between Alice and Bob

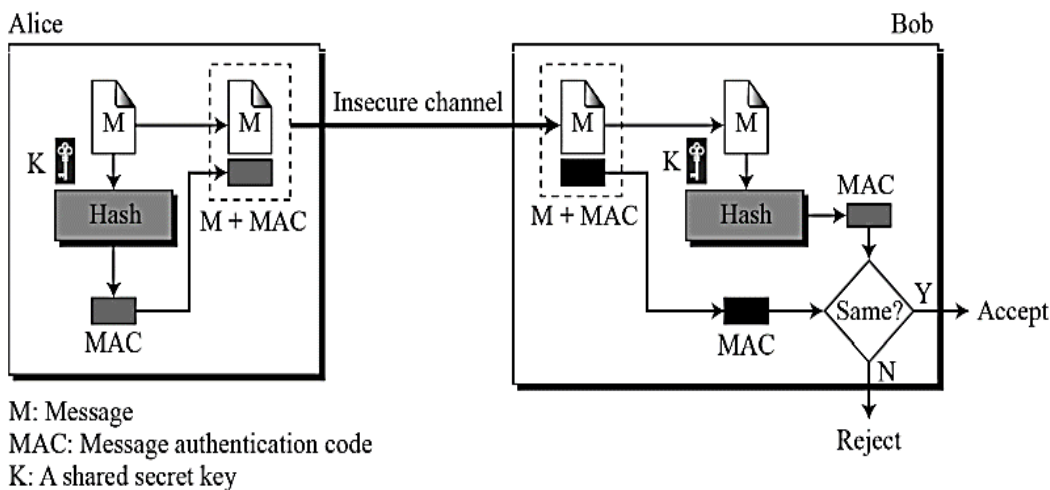


Figure: Message Authentication Code

MAC Security

How can Eve forge a message without having the key?

1. If size of the key allows exhaustive search, Eve may try all possible keys to digest the message.
2. Use preimage attack.
3. Given some pairs of messages and their MACs, Eve can manipulate them to come up with a new message and its digest.

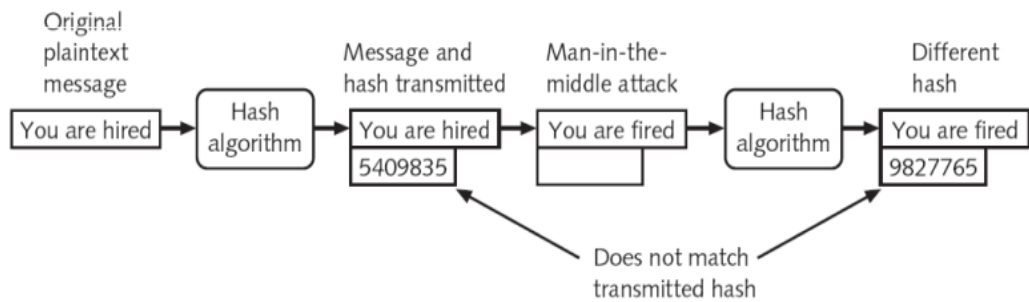
It is important to note that security of a MAC depends on the security of the underlying hash algorithm.

Nested MAC

To improve MAC security, nested MACs were designed in which hashing is performed twice.

Message Integrity - Hash Functions

A cryptographic hash function takes a message of arbitrary length and creates a message digest of fixed length, also called hash. Hashing cryptography is used to provide integrity, to ensure that information has not been altered by an unauthorized person or malicious software. It is important to note that security of a MAC depends on the security of the underlying hash algorithm.



It is important to note that the message and its digest are sent to the receiver to verify the integrity.

A hash function takes an input, m , and computes a fixed-size string $H(m)$. A cryptographic hash function also needs to satisfy an additional property:

It is computationally infeasible to find any two different messages x and y such that $H(x) = H(y)$.

Message Integrity – Hash Functions

Common hash algorithms are MD5 [RFC 1321] and SHA-1 [FIPS 1995].

After the cryptographic hash function. A naive step to perform message integrity would be:

- **Alice creates message m and send Bob $(m, H(m))$.**
- **After Bob receives $(m, H(m))$, he uses the same hash function to check whether after hashing m , the result is equal to the $H(m)$ he received from Alice.**

The approach is flawed because other party can also send message to bob $(m', H(m'))$ claiming that this is from Alice, and Bob has no way to tell even the hash result is correct. In order to prove the message is indeed from Alice, Alice and Bob should have a shared secrets. This shared secret is called authentication key. Using this shared secret, message integrity can be performed as:

- 1) Alice creates message m , concatenates s to m , calculates $H(m+s)$, sends Bob $(m, H(m+s))$. $H(m+s)$ here is called the message authentication code (MAC).
- 2) Bob receives $(m, H(m+s))$, calculates the MAC using his s , check whether it is the same $H(m+s)$ he received from the message.



Message integrity does not imply or require encrypting the message between Alice and Bob.

14.6 Email

Email is one of the most widely used and regarded network services. Currently message contents are not secure. They may be inspected either in transit or by suitably privileged users on destination system.

Email Security

Email security is the term for any procedure that protects email content and accounts against unauthorized access. Email service providers have email security measures in place to secure client accounts and information from hackers. Such measures include:

- email servers with strong password and access control mechanisms;
- encrypted email messages (both inboxed or in transit);
- web application firewalls; and
- spam filtering software.

Email is popular with hackers as a tool for spreading malware, spam, and phishing attacks. They use deceptive messages to trick recipients into sharing sensitive information, resulting in identity theft. The attackers lure people into opening attachments or clicking hyperlinks that install

malware (such as email viruses) on the user's device. Email is also a main entry point for attackers looking to access an enterprise network and breach valuable company data. Despite the fact that many of us can't function without checking our email on a regular basis, we often take the privacy and security of our inboxes for granted. Email is a prime target for hackers and data thieves — and it's not a particularly difficult one, either. And then, there's the issue of surveillance. While most of us aren't spies emailing the country's secrets to unfriendly nations, the idea that somebody might have a back door into our personal emails is more than a little unsettling. Of course, that doesn't mean that secure, private email isn't possible. It's just up to you to take a few precautions to keep your email safe:

1) Use two-factor authentication

The basic principle of two-factor authentication is simple: combine something you know with something you have. One example is a debit card, which requires you to have both your physical card and your PIN to verify your identity. By enabling two-factor authentication (or two-step verification), you aren't putting all of your faith in a password. That's a good thing, considering how weak many of our passwords are. For Gmail, setting up two-step verification is as simple as clicking a button and entering in your mobile number. For Windows Mail, or Outlook, it's a similar process. Just log in, go to your "Password and security" tab and click "Set up two-step verification." Now that you've enabled two-factor authentication, a hacker with your password is out of luck — unless they've also managed to steal your cell phone.

2) Limit forwarding

When we're sent a message we want to share, we often click "Forward" without thinking about the consequences.

- Where is the message going?
- Who will see it?
- Where will it be stored?

If your email is hosted on a corporate server, it is likely there are certain security measures in place to protect any sensitive information contained in your private email. When someone forwards an internal email to a recipient outside of your company. However, you are exposing that data (as well as any other emails in the forwarded chain) to potentially unsecured, unencrypted servers. Similarly, if you're a covered entity sending email containing protected health information (PHI) to a business associate, all it takes is one employee to forward that email to an unauthorized recipient to violate HIPAA.

3) Set Expiration Dates on Your Messages

While some of us can't stand a messy inbox, the average user doesn't bother cleaning up their private email, often seeing deleting email as a waste of time. Considering more than 50 percent of us receive at least 11 emails a day, can you blame them? That means that any sensitive information you send to a client could very well be sitting there months later. At that point, you no longer control the fate of your data. Luckily, Virtru lets you set There should be an expiration date on your email, so that after a certain date, it will no longer be readable by the recipient (or anyone else, for that matter).

4) Understand your service provider's TOS

Your email provider's terms of service can tell you a lot more than their media interviews and advertisements can.

- For starters, it will let you know what kind of security they are offering you.
- are they encrypting messages on their server?
- do they have protections against brute-force attacks?
- is there any guarantee that your data is being protected?

While you might think your email provider has your best interests in mind, there's a good chance that they don't have the same expectations you do. Take Google for example, which openly passes private email through automated scanning. After reading your email provider's TOS, you'll likely realize that keeping your private email secure isn't their first priority — that's entirely up to you.

5) Encrypt your email

The best way to keep your private email away from prying eyes and hackers is to use encryption. Encryption protects your private email by jumbling up your messages, making

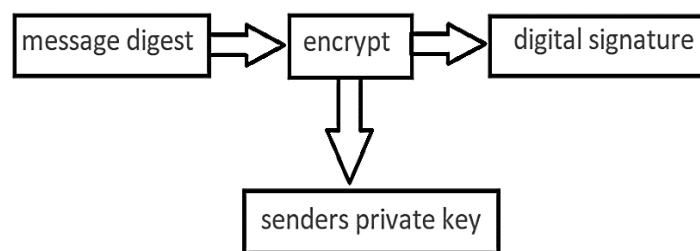
them impossible to decipher unless you explicitly authorize someone to read them. If you are using a client-side encryption service like Virtru, even if your inbox is compromised, the contents of your message will be unreadable. Likewise, you don't have to worry about your messages being intercepted after you send them, either by hackers or nosey service providers. As an added bonus, if your email ends up getting stored on a server outside of your control, you still have power over who gets to see it – and you can revoke that permission at any time. While email may not have been designed to be secure, but users can enjoy added privacy and security with a few workarounds. Virtru works with the email service you're already using to provide true client-side email encryption for your messages and attachments.

Privacy Enhanced Mail

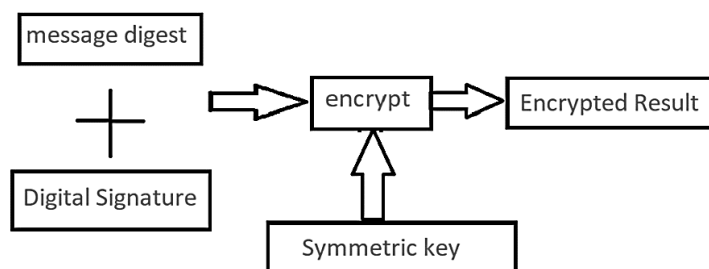
PEM adopted by the Internet Architecture Board (IAB) to provide secure electronic mail communication over the internet.

Steps of Privacy Enhanced Mail (PEM)

- 1) **Canonical Conversions** - Privacy Enhanced Mail (PEM) transforms each email message into an abstract canonical representation. This means that regardless of the architecture and the operating system of the sending and receiving computers. The email message always travels in a uniform, independent format.
- 2) **Digital Signature** - In this step, the digital signature is generated by encrypting the message digest of an email message with the sender's private key.



- 3) **Encryption** - The encrypted message is generated by encrypting the original message and digital signature together along with the symmetric key as shown in the figure below. This step is very crucial in order to obtain the confidentiality.



- 4) **Base 64encoding** - This is the last step where the binary output is transformed into character output. The binary output which is 24 bits is divided into 4 equal sets and mapped with the 8-bit character output generating a decimal code. Now PEM uses a separate map table and each number from the code generated is mapped with its corresponding value from the mapping table and binary equivalent corresponding to the 8-bit ASCII of the character is written.

Printable Encoding Characters

Privacy Enhanced Mail (PEM) transforms each email message into an abstract canonical representation. This means that regardless of the architecture and the operating system of the sending and receiving computers. The email message always travels in a uniform, independent format.

Pretty Good Privacy (PGP)

PGP provides a confidentiality and authentication service that can be used for electronic mail and file storage applications. A number of reasons can be cited for this growth. The prominent reasons are:

- available free worldwide
- It is based on extremely secure algorithm.
- wide range of applicability
- not developed by governmental organization.

Operational Description (PGP)

The actual operation of PGP, consists of five services:

- 1) **Authentication** - The basic steps of authentication can be clearly understood with the help of figure

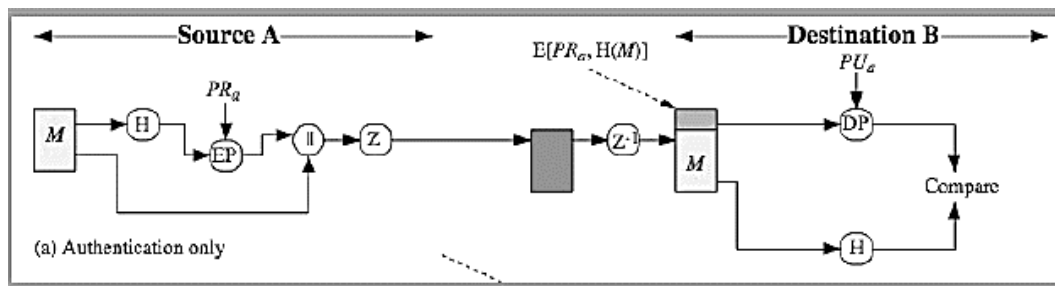


Figure: Authentication

In the figure, it can be clearly seen

- sender creates message
- make SHA-1160-bit hash of message
- attached RSA signed hash to message
- receiver decrypts & recovers hash code
- receiver verifies received message hash

2) Confidentiality & Authentication

1. can use both services on same message
2. create signature & attach to message
3. encrypt both message & signature
4. attach RSA/ElGamal encrypted session key

- 3) **Compression** - By default, PGP compresses message after signing but before encrypting. So can store uncompressed message & signature for later verification & because compression is non-deterministic. It uses ZIP compression algorithm

- 4) **Email Compatibility** - When using PGP will have binary data to send (encrypted message etc). However email was designed only for text. Hence PGP must encode raw binary data into printable ASCII characters. It uses radix-64 algorithm. It then maps 3 bytes to 4 printable chars. Later it also appends a CRC. PGP also segments messages if too big.

Signed Mail

In a signed mail, a user writes the message as clear text. The message digest is being calculated (using SHA-1 or MD5). The message digest is being encrypted using the signer's private key (DSS or RSA).

Encrypted mail

When a user writes an encrypted mail, he writes the message as clear-text. A random session key is being created (tripleDES or RC2). The message is being encrypted using the random session key. For every recipient, the session key is being encrypted using the recipient's public key (DH or RSA).

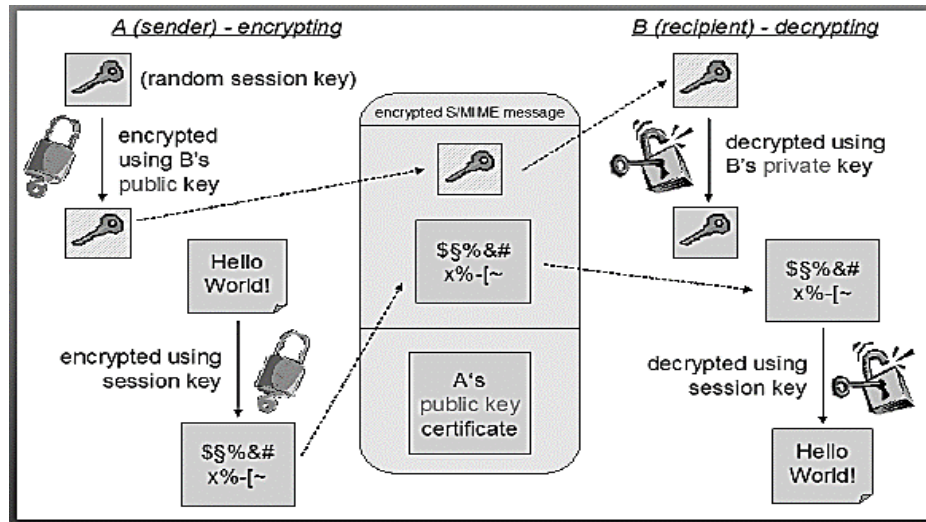


Figure: Encrypted mail

S/MIME Cryptographic Algorithms

- digital signatures: DSS & RSA
- hash functions: SHA-1 & MD5
- session key encryption: ElGamal & RSA
- message encryption: AES, Triple-DES, RC2/40 and others
- MAC: HMAC with SHA-1

S/MIME Functions

- Enveloped data
encrypted content and associated keys
- Signed data
encoded message + signed digest
- Clear-signed data
cleartext message + encoded signed digest
- Signed & enveloped data
nesting of signed & encrypted entities

E-mail Hacking

Email hacking can be done in any of the following ways:

- 1) **E-mail Spam** - E-mail spamming is an act of sending Unsolicited Bulk E-mails (UBI) which one has not asked for. Spams are the junk mails sent by commercial companies as an advertisement of their products and services.
- 2) **Virus** - Some emails may incorporate with files containing malicious script which when run on your computer may lead to destroy your important data.
- 3) **Email Phishing** - Email phishing is an activity of sending emails to a user claiming to be a legitimate enterprise. Its main purpose is to steal sensitive information such as usernames, passwords, and credit card details. Such emails contain links to websites that are infected with malware and direct the user to enter details at a fake website whose look and feels are same to legitimate one.

- 4) **Email Spamming and Junk Mails** - Email spamming is an act of sending Unsolicited Bulk E-mails (UBI) which one has not asked for. Email spams are the junk mails sent by commercial companies as an advertisement of their products and services.

Email Spamming Problems

Spams may cause the following problems:

- It floods your e-mail account with unwanted e-mails, which may result in loss of important e-mails if inbox is full.
- Time and energy is wasted in reviewing and deleting junk emails or spams.
- It consumes the bandwidth that slows the speed with which mails are delivered.
- Some unsolicited email may contain virus that can cause harm to your computer.

Blocking Spams

Following ways will help you to reduce spams:

- a) While posting letters to newsgroups or mailing list, use a separate e-mail address than the one you used for your personal e-mails.
- b) Don't give your email address on the websites as it can easily be spammed.
- c) Avoid replying to emails which you have received from unknown persons.
- d) Never buy anything in response to a spam that advertises a product.

E-mail Cleanup and Archiving

In order to have light weighted Inbox, it's good to archive your inbox from time to time. The steps to clean up and archive your Outlook inbox are:

- a) Select File tab on the mail pane.
- b) Select Cleanup Tools button on account information screen.
- c) Select Archive from cleanup tools drop down menu.
- d) Select Archive this folder and all subfolders option and then click on the folder that you want to archive. Select the date from the Archive items older than: list. Click Browse to create new .pst file name and location. Click OK.

14.7 Firewalls

A firewall can be defined as a special type of network security device or a software program that monitors and filters incoming and outgoing network traffic based on a defined set of security rules. It acts as a barrier between internal private networks and external sources (such as the public Internet).

A firewall is a hardware, software or a combination of both that monitors and filters traffic packets that attempt to either enter or leave the protected private network. It is a tool that separates a protected network or part of a network, and now increasingly a user PC, from an unprotected network – the “bad network” like the Internet.

Firewall Purpose

The primary purpose of a firewall is to allow non-threatening traffic and prevent malicious or unwanted data traffic for protecting the computer from viruses and attacks. A firewall is a cybersecurity tool that filters network traffic and helps users block malicious software from accessing the Internet in infected computers.

Basic Security Functions of Firewalls

Most firewalls perform two basic security functions:

- 1) Packet filtering based on accept or deny policy that is itself based on rules of the security policy.
- 2) Application proxy gateways that provide services to the inside users and at the same time protect each individual host from the “bad” outside users.

Firewalls Security Policies

These policies are consolidated into two commonly used firewall security policies:

- Deny-everything-not-specifically-allowed which sets the firewall in such a way that it denies, all traffic and services except a few that are added as the organizations needs develop.
- Allow-everything-not-specifically-denied which lets in all the traffic and services except those on the “forbidden” list which is developed as the organization’s dislikes grow.

Risks of Not Having a Firewall

Some of the important risks of not having a firewall are:

- 1) **Open Access** - If a computer is running without a firewall, it is giving open access to other networks. This means that it is accepting every kind of connection that comes through someone. In this case, it is not possible to detect threats or attacks coming through our network. Without a firewall, we make our devices vulnerable to malicious users and other unwanted sources.
- 2) **Lost or Comprised Data** - Without a firewall, we are leaving our devices accessible to everyone. This means that anyone can access our device and have complete control over it, including the network. Risks of Not Having a Firewall. In this case, cybercriminals can easily delete our data or use our personal information for their benefit.
- 3) **Network Crashes** - In the absence of a firewall, anyone could access our network and shut it down. It may lead us to invest our valuable time and money to get our network working again. Therefore, it is essential to use firewalls and keep our network, computer, and data safe and secure from unwanted sources.

How does a firewall work?

A firewall system analyzes network traffic based on pre-defined rules. It then filters the traffic and prevents any such traffic coming from unreliable or suspicious sources. It only allows incoming traffic that is configured to accept.

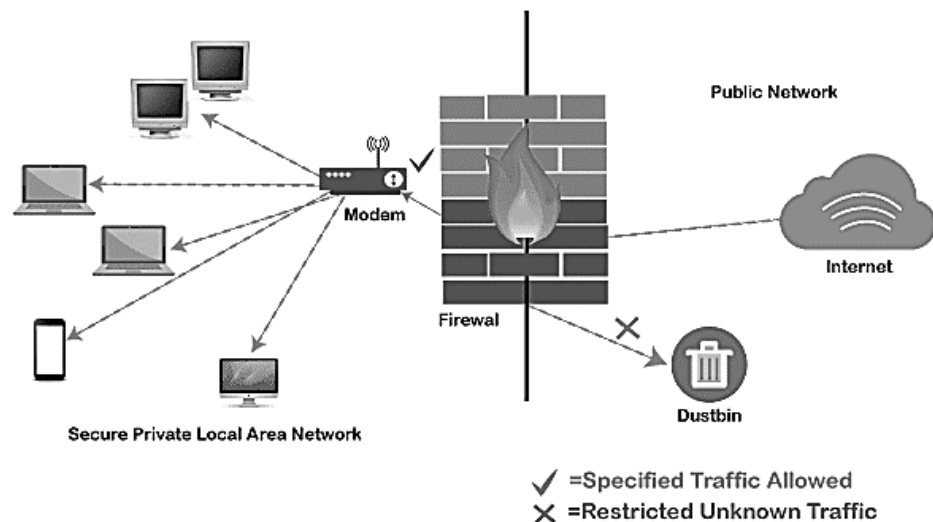


Figure: Working of a Firewall

Typically, firewalls intercept network traffic at a computer's entry point, known as a port. Firewalls perform this task by allowing or blocking specific data packets (units of communication transferred over a digital network) based on pre-defined security rules. Incoming traffic is allowed only through trusted IP addresses, or sources.

Functions of Firewall

As stated above, the firewall works as a gatekeeper. It analyzes every attempt coming to gain access to our operating system and prevents traffic from unwanted or non-recognized sources. Since the firewall acts as a barrier or filter between the computer system and other networks (i.e., the public Internet), we can consider it as a traffic controller. Therefore, a firewall's primary function is to

secure our network and information by controlling network traffic, preventing unwanted incoming network traffic, and validating access by assessing network traffic for malicious things such as hackers and malware. Generally, most operating systems (for example - Windows OS) and security software come with built-in firewall support. Therefore, it is a good idea to ensure that those options are turned on. Additionally, we can configure the security settings of the system to be automatically updated whenever available. Firewalls have become so powerful, and include a variety of functions and capabilities with built-in features like Network Threat Prevention, Application and Identity-Based Control, Hybrid Cloud Support, Scalable Performance, Network Traffic Management and Control, Access Validation and Record and Report on Events.

Limitations of Firewall

When it comes to network security, firewalls are considered the first line of defense. But the question is whether these firewalls are strong enough to make our devices safe from cyber-attacks. The answer may be "no". The best practice is to use a firewall system when using the Internet. However, it is important to use other defense systems to help protect the network and data stored on the computer. Because cyber threats are continually evolving, a firewall should not be the only consideration for protecting the home network. The importance of using firewalls as a security system is obvious; however, firewalls have some limitations:

- Firewalls cannot stop users from accessing malicious websites, making it vulnerable to internal threats or attacks.
- Firewalls cannot protect against the transfer of virus-infected files or software.
- Firewalls cannot prevent misuse of passwords.
- Firewalls cannot protect if security rules are misconfigured.
- Firewalls cannot protect against non-technical security risks, such as social engineering.
- Firewalls cannot stop or prevent attackers with modems from dialing in to or out of the internal network.
- Firewalls cannot secure the system which is already infected.

Therefore, it is recommended to keep all Internet-enabled devices updated. This includes the latest operating systems, web browsers, applications, and other security software (such as anti-virus). Besides, the security of wireless routers should be another practice. The process of protecting a router may include options such as repeatedly changing the router's name and password, reviewing security settings, and creating a guest network for visitors.

Types of Firewall

Depending on their structure and functionality, there are different types of firewalls. The following is a list of some common types of firewalls:

1) Packet-filtering Firewalls

A packet filtering firewall is the most basic type of firewall. It acts like a management program that monitors network traffic and filters incoming packets based on configured security rules. These firewalls are designed to block network traffic IP protocols, an IP address, and a port number if a data packet does not match the established rule-set. While packet-filtering firewalls can be considered a fast solution without many resource requirements, they also have some limitations. Because these types of firewalls do not prevent web-based attacks, they are not the safest.

2) Circuit-level Gateways

Circuit-level gateways are another simplified type of firewall that can be easily configured to allow or block traffic without consuming significant computing resources. These types of firewalls typically operate at the session-level of the OSI model by verifying TCP (Transmission Control Protocol) connections and sessions. Circuit-level gateways are designed to ensure that the established sessions are protected. Typically, circuit-level firewalls are implemented as security software or pre-existing firewalls. Like packet-filtering firewalls, these firewalls do not check for actual data, although they inspect information about transactions. Therefore, if a data contains malware, but follows the correct TCP connection, it will pass through the gateway. That is why circuit-level gateways are not considered safe enough to protect our systems.

3) Application-level Gateways (Proxy Firewalls)

Proxy firewalls operate at the application layer as an intermediate device to filter incoming traffic between two end systems (e.g., network and traffic systems). That is why these firewalls are called Application-level Gateways. This protects the client's identity and other suspicious information, keeping the network safe from potential attacks. Once the connection is established, the proxy firewall inspects data packets coming from the source. If the contents of the incoming data packet are protected, the proxy firewall transfers it to the client. This approach creates an additional layer of security between the client and many different sources on the network.

4) Stateful Multi-layer Inspection (SMLI) Firewalls

Stateful multi-layer inspection firewalls include both packet inspection technology and TCP handshake verification, making SMLI firewalls superior to packet-filtering firewalls or circuit-level gateways. Additionally, these types of firewalls keep track of the status of established connections. In simple words, when a user establishes a connection and requests data, the SMLI firewall creates a database (state table). The database is used to store session information such as source IP address, port number, destination IP address, destination port number, etc. Connection information is stored for each session in the state table. Using stateful inspection technology, these firewalls create security rules to allow anticipated traffic. In most cases, SMLI firewalls are implemented as additional security levels. These types of firewalls implement more checks and are considered more secure than stateless firewalls. This is why stateful packet inspection is implemented along with many other firewalls to track statistics for all internal traffic. Doing so increases the load and puts more pressure on computing resources. This can give rise to a slower transfer rate for data packets than other solutions.

5) Next-generation Firewalls (NGFW)

Many of the latest released firewalls are usually defined as *next-generation firewalls*. However, there is no specific definition for next-generation firewalls. This type of firewall is usually defined as a security device combining the features and functionalities of other firewalls. These firewalls include deep-packet inspection (DPI), surface-level packet inspection, and TCP handshake testing, etc. NGFW includes higher levels of security than packet-filtering and stateful inspection firewalls. Unlike traditional firewalls, NGFW monitors the entire transaction of data, including packet headers, packet contents, and sources. NGFWs are designed in such a way that they can prevent more sophisticated and evolving security threats such as malware attacks, external threats, and advance intrusion.

6) Threat-focused NGFW

Threat-focused NGFW includes all the features of a traditional NGFW. Additionally, they also provide advanced threat detection and remediation. These types of firewalls are capable of reacting against attacks quickly. With intelligent security automation, threat-focused NGFW set security rules and policies, further increasing the security of the overall defense system. In addition, these firewalls use retrospective security systems to monitor suspicious activities continuously. They keep analyzing the behavior of every activity even after the initial inspection. Due to this functionality, threat-focus NGFW dramatically reduces the overall time taken from threat detection to cleanup.

7) Network Address Translation (NAT) Firewalls

Network address translation or NAT firewalls are primarily designed to access Internet traffic and block all unwanted connections. These types of firewalls usually hide the IP addresses of our devices, making it safe from attackers. When multiple devices are used to connect to the Internet, NAT firewalls create a unique IP address and hide individual devices' IP addresses. As a result, a single IP address is used for all devices. By doing this, NAT firewalls secure independent network addresses from attackers scanning a network for accessing IP addresses. This results in enhanced protection against suspicious activities and attacks. In general, NAT firewalls work similarly to proxy firewalls.

Like proxy firewalls, NAT firewalls also work as an intermediate device between a group of computers and external traffic.

8) Cloud Firewalls

Whenever a firewall is designed using a cloud solution, it is known as a cloud firewall or FaaS (firewall-as-a-service). Cloud firewalls are typically maintained and run on the Internet by third-party vendors. This type of firewall is considered similar to a proxy firewall. The reason for this is the use of cloud firewalls as proxy servers. However, they are configured based on requirements. The most significant advantage of cloud firewalls is scalability. Because cloud

firewalls have no physical resources, they are easy to scale according to the organization's demand or traffic-load. If demand increases, additional capacity can be added to the cloud server to filter out the additional traffic load. Most organizations use cloud firewalls to secure their internal networks or entire cloud infrastructure.

9) Unified Threat Management (UTM) Firewalls

UTM firewalls are a special type of device that includes features of a stateful inspection firewall with anti-virus and intrusion prevention support. Such firewalls are designed to provide simplicity and ease of use. These firewalls can also add many other services, such as cloud management, etc.

Which Firewalls Architecture is Best?

When it comes to selecting the best firewall architecture, there is no need to be explicit. It is always better to use a combination of different firewalls to add multiple layers of protection. For example, one can implement a hardware or cloud firewall at the perimeter of the network, and then further add individual software firewall with every network asset.

Besides, the selection usually depends on the requirements of any organization. However, there are a few factors which can be considered for the right selection of firewall. These factors are:

- 1) **Size of the organization** - If an organization is large and maintains a large internal network, it is better to implement such firewall architecture, which can monitor the entire internal network.
- 2) **Availability of resources** - If an organization has the resources and can afford a separate firewall for each hardware piece, this is a good option. Besides, a cloud firewall may be another consideration.
- 3) **Requirement of multi-level protection** - The number and type of firewalls typically depend on the security measures that an internal network requires. This means, if an organization maintains sensitive data, it is better to implement multi-level protection of firewalls. This will ensure data security from hackers.

Firewalls v/s Anti-Viruses

Attributes	Firewall	Anti-virus
Definition	A firewall is defined as the system which analyzes and filters incoming or outgoing data packets based on pre-defined rules.	Anti-virus is defined as the special type of software that acts as a cyber-security mechanism. The primary function of Anti-virus is to monitor, detect, and remove any apprehensive or distrustful file or software from the device.
Structure	Firewalls can be hardware and software both. The router is an example of a physical firewall, and a simple firewall program on the system is an example of a software firewall.	Anti-virus can only be used as software. Anti-virus is a program that is installed on the device, just like the other programs.
Implementation	Because firewalls come in the form of hardware and software, a firewall can be implemented either way.	Because Anti-virus comes in the form of software, therefore, Anti-virus can be implemented only at the software level. There is no possibility of implementing Anti-virus at the hardware level.
Responsibility	A firewall is usually defined as a network controlling system. It means that firewalls are primarily responsible for monitoring and filtering network traffic.	Anti-viruses are primarily responsible for detecting and removing viruses from computer systems or other devices. These viruses can be in the form of infected files or software.
Scalability	Because the firewall supports both types of implementations, hardware, and software, therefore, it is more scalable than anti-virus.	Anti-viruses are generally considered less scalable than firewalls. This is because anti-virus can only be implemented at the software level. They don't support hardware-level implementation.

Threats	A firewall is mainly used to prevent network related attacks. It mainly includes external network threats? For example- Routing attacks and IP Spoofing.	Anti-virus is mainly used to scan, find, and remove viruses, malware, and Trojans, which can harm system files and software and share personal information (such as login credentials, credit card details, etc.) with hackers.
---------	--	---

Summary

Data on the network is not secret and therefore requires to be kept secure from undesirable persons sitting behind the machines attached to the network.

The malicious intentions may include bringing down of servers attached to the network, using people's private information like credit card numbers for fraudulent activities and sabotaging of major organizations by accessing their websites. It is therefore aimed to secure data and prevent from eavesdroppers from listening to and stealing data. The user data on a computer is also protected by providing password restricted access to the data and resources so that only authorized people get to use these. Security aspects also involve identifying miscreants and thwarting their attempts to cause damage to the network among other resources.

Authentication involves verifying of the antecedents of the person who has requested for services from a remote machine or access to the remote machine either through physically or by sending an e-mail before allowing him or her to do so. Authentication involves a process to authenticate person's identity to a remote machine.

Integrity involves the veracity of the message which is received by a remote machine. In other words, it is indeed the same message without any alteration which was sent by the source machine. In this case, cyclic redundancy code method will not be enough as intruders in the system or communication channel may deliberately alter the message. Security should ensure that nobody along the entire route should be able to alter the message.

Confidentiality: It ensures that no person should be able to read the message on the way. This necessitates implementation of the encryption techniques down the line.

The message is encrypted at the sender end and decrypted at the receiving end to maintain privacy with the help of the encryption and decryption techniques. The secret key and public key techniques are the available techniques with their advantages and disadvantages.

Substitution and transposition ciphers are two categories of ciphers used in classical cryptography. Substitution and transposition differ in how chunks of the message are handled by the encryption process.

Keywords

Ciphertext: This is the encrypted message generated by applying the algorithm to the plaintext message using the secret key.

IP-spoofing: Like honeypots, IP spoofing involves the interception of data packets by a computer successfully pretending to be a trusted server/resource.

Maliciously: Coded Websites - Maliciously coded Websites create charitable websites enabling a user to make donations and thus stealing the vital personal information.

Packet Sniffers: Packet sniffers are the technique used to capture data streams over a network to obtain sensitive data like usernames, passwords, credit card numbers, etc.

Password Attacks: A 'Password Attack' includes a number of techniques used by hackers to steal passwords.

Phishing: Emails with titles such as, "URGENT: Update Account Status" are all attempts by a spammer to "phish" the account details.

Plaintext: It is the text message to be transmitted on which an algorithm is applied.

Private Key: The key of a key pair, which is used to create a digital signature. It can be used to sign a message that only the corresponding public key can verify.

Public Key: It is the key of a key pair that is used to verify a digital signature. Key pair consists of private and public key.

Secret Key: They constitute a part of algorithm for encryption and decryption of the message.

Self Assessment

1. In secret key encryption, the secret key is used for
 - A. Encryption only
 - B. Encryption and decryption
 - C. Decrypting the encrypted message
 - D. None of the given choices

2. In the public key encryption, the public key is used for of the message.
 - A. Encryption
 - B. Decryption
 - C. Encryption and Decryption
 - D. None of the given choices

3. Encryption and decryption normally takes care of of a network.
 - A. Consistency
 - B. security
 - C. authentication
 - D. privacy

4. In public key encryption the private is used to the message to the plaintext.
 - A. Encrypt
 - B. Encryption and Decryption
 - C. Decrypt
 - D. None of the given choices

5. involves the interception of data packets by a computer successfully pretending to be a trusted server/resource.
 - A. IP spoofing
 - B. Brute force attack
 - C. Sniffing
 - D. Phishing

6. An Asymmetric key cipher uses
 - A. 1 key
 - B. 2 key
 - C. 3 key
 - D. 4 key

7. The shift cipher can also be referred as:
 - A. Caesar Cipher
 - B. Vigenere Cipher
 - C. RSA
 - D. None of the given choices

8. The cryptographic Algorithms or the ciphers are divided into:

- A. Two groups
 - B. Four groups
 - C. One single group
 - D. Zero single groups
9. In symmetric key cryptography, the key used by the sender and the receiver is:
- A. Different
 - B. Shared
 - C. Two keys are used
 - D. Same keys are used
10. Cryptography is a word with Greek origin which means:
- A. Corrupting data
 - B. Secret writing
 - C. Open writing
 - D. Closed writing
11. In cryptography when we treat the text at the bit level, then every character is replaced by:
- A. 4 bits
 - B. 6 bits
 - C. 8 bits
 - D. 10 bits
12. Modern cryptanalysis makes simple substitution and transposition ciphers obsolete.
- A. True
 - B. False
13. According to the unicity distance of English, 20 letters of ciphertext are required to crack a mixed alphabet simple substitution.
- A. True
 - B. False
14. In substitution cipher, a bijective function is used on the characters' positions to encrypt and an inverse function to decrypt.
- A. True
 - B. False
15. Traditionally, the ciphertext is written out in blocks of fixed length, omitting punctuation and spaces.
- A. True
 - B. False

Answer for Self Assessment

- | | | | | |
|-------|-------|-------|-------|-------|
| 1. B | 2. A | 3. D | 4. C | 5. A |
| 6. B | 7. A | 8. A | 9. B | 10. B |
| 11. C | 12. A | 13. B | 14. A | 15. A |

Review Questions

- 1) What are different criterions to keep information private when it is sent over a public network?
- 2) How does the encryption affect performance of network?
- 3) There are certain information bases on the Internet that need to be prevented by undesirable person to get. How can undesirable person be kept from accessing this?
- 4) How do we keep our own and other people's computers safe from hackers? Explain with the help of a hypothetical situation.
- 5) What is a Cipher? Why are cipher used for large messages?
- 6) Describe briefly two kinds of security attacks, which can be directed against an Internet connected computer system.
- 7) What is the difference between secret key and public key encryption?
- 8) What is cryptography? What are the benefits of using this technique?
- 9) What do you mean by substitution and transposition ciphers? Differentiate between the two.

**Further Readings**

Books Andrew S. Tanenbaum, *Computer Networks*, Prentice Hall
Behrouz A. Forouzan and Sophia Chung Fegan, *Data Communications and Networking*, McGraw-Hill Companies
Burton, Bill, *Remote Access for Cisco Networks*, McGraw-Hill, Osborne Media
Dale Tesch/Greg Abelar, *Security Threat Mitigation and Response: Understanding CS-MARS*, Cisco Press, Sep. 26, 2006.
Gary Halleen/Greg Kellogg, *Security Monitoring with Cisco Security MARS*, Cisco Press, Jul. 6, 2007

LOVELY PROFESSIONAL UNIVERSITY

Jalandhar-Delhi G.T. Road (NH-1)

Phagwara, Punjab (India)-144411

For Enquiry: +91-1824-521360

Fax.: +91-1824-506111

Email: odl@lpu.co.in