Module 1 Computer Networks

Introduction to Networks: Network hardware - Network software - Reference models
Physical Layer: Guided transmission media, Wireless transmission

Uses of Computer Networks

i. Business Application – resource sharing, Virtual Private Networks, Client server computing, Web applications, VoIP, Desktop Sharing, e-commerce
ii. Home Applications – connect Internet
iii. Mobile users
iv. Social issues

NETWORK HARDWARE

Computer networks fit, but two dimensions stand out as important:

1. Transmission technology and
2. Scale.

Two types of transmission technology that are in widespread use:

i. Broadcast links - the communication channel is shared by all the machines on the network; packets sent by any machine are received by all the others - known as multicasting.
ii. Point-to-point links - Point-to-point links connect individual pairs of machines – known as unicasting.

To go from the source to the destination on a network made up of packets (short messages).

An alternative criterion for classifying networks is by scale. Distance is important as a classification metric because different technologies are used at different scales.

Interprocessor distance	Processors located in same	Example
1 m	Square meter	Personal area network
10 m	Room	Local area network
100 m	Building	
1 km	Campus	
10 km	City	Metropolitan area network
100 km	Country	Wide area network
1000 km	Continent	
10,000 km	Planet	The Internet

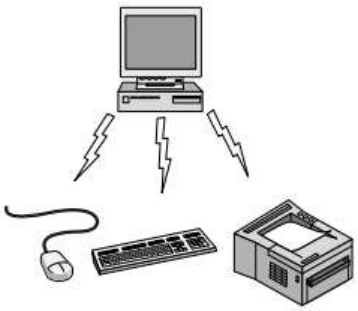
Classification of interconnected processors by scale.

1. PAN - Personal Area Networks
2. LAN - Local Area Networks
3. MAN - Metropolitan Area Networks
4. WAN - Wide Area Networks
5. Internetworks or Internet

1. PAN - Personal Area Networks

- PANs (Personal Area Networks) let devices communicate over the range of a person through wired or wireless.

- Almost every computer has an attached monitor, keyboard, mouse, and printer - this connection must be done with cables.
- A common example is a wireless network that connects a computer with its peripherals called Bluetooth to connect these components without wires.
- PANs can also be built with other technologies that communicate over short ranges, such as RFID on smartcards and library books.



Bluetooth PAN configuration.

2. LAN - Local Area Networks

- A LAN is a privately owned network that operates within and nearby a single building like a home, office or factory.
- LANs are widely used to connect personal computers and consumer electronics to let them share resources (e.g., printers) and exchange information.
- When **LANs (Wired)** are used by companies, they are called enterprise networks.
- **Wireless LANs** are very popular these days, especially in homes, older office buildings, cafeterias, and other places where it is too much trouble to install cables.

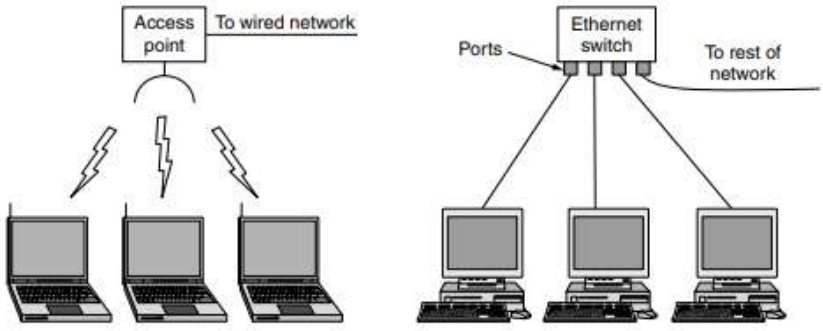
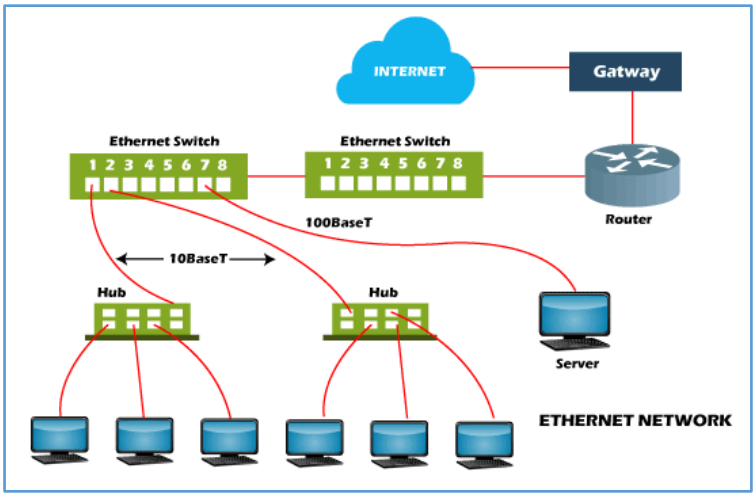


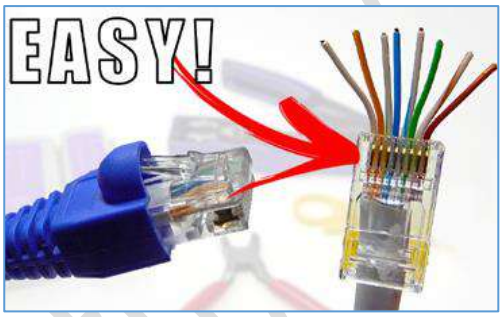
Figure Wireless and wired LANs. (a) 802.11. (b) Switched Ethernet.

- Each computer talks to a device in the ceiling as shown in Fig. 1-8(a). This device, called an AP (Access Point), wireless router, or base station, relays packets between the wireless computers and also between them and the Internet.
- There is a standard for wireless LANs called IEEE 802.11, popularly known as WiFi, which has become very widespread - It runs at speeds anywhere from 11 to hundreds of Mbps.
- Wired LANs use a range of different transmission technologies - Most of them use copper wires, but some use optical fiber.

- LANs are restricted in size, which means that the worst-case transmission time is bounded and known in advance. Knowing these bounds helps with the task of designing network protocols.
- Typically, wired LANs run at speeds of 100 Mbps to 1 Gbps, have low delay (microseconds or nanoseconds), and make very few errors. Newer LANs can operate at up to 10 Gbps.

- Compared to wireless networks, wired LANs exceed them in all dimensions of performance. It is just easier to send signals over a wire or through a fiber than through the air.
- The topology of many wired LANs is built from point-to-point links. IEEE 802.3, popularly called **Ethernet**, is, by far, the most common type of wired LAN.

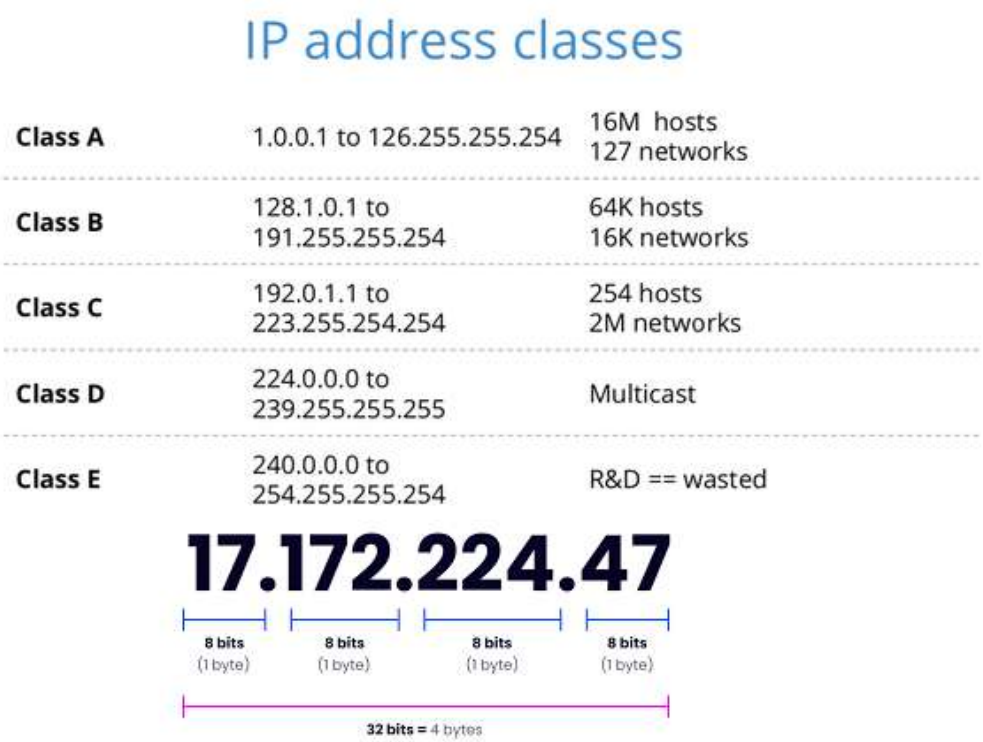




- Each computer speaks the Ethernet protocol and connects to a box called a switch with a point-to-point link. Hence the name.

- A switch has multiple ports, each of which can connect to one computer.
- The job of the switch is to relay packets between computers that are attached to it, using the address in each packet to determine which computer to send it to.
- To build larger LANs, switches can be plugged into each other using their ports.
- It is also possible to divide one large physical LAN into two smaller logical LANs (Virtual LAN or VLAN).
- **Both wireless and wired broadcast networks** can be **divided into static and dynamic designs**, depending on how the channel is allocated.
- A typical static allocation would be to divide time into discrete intervals and use a round-robin algorithm, allowing each machine to broadcast only when its time slot comes up.
- Dynamic allocation methods for a common channel are either centralized or decentralized.
- In the centralized channel allocation method, there is a single entity, for example, the base station in cellular networks, which determines who goes next.

- In the decentralized channel allocation method, there is no central entity; each machine must decide for itself whether to transmit. You might think that this approach would lead to confusion, but it does not.



- While we could think of the home network as just another LAN, it is more likely to have different properties than other networks.
    - First, the networked devices have to be very easy to install.
    - Second, the network and devices have to be fool proof in operation.
    - Third, low price is essential for success.
    - Fourth, it must be possible to start out with one or two devices and expand the reach of the network gradually. This means no format wars.
    - Fifth, security and reliability will be very important.

3. MAN - Metropolitan Area Networks

- A MAN (Metropolitan Area Network) covers a city.
- The best-known examples of MANs are the cable television networks available in many cities.
- These systems grew from earlier community antenna systems used in areas with poor over-the-air television reception.
- In those early systems, a large antenna was placed on top of a nearby hill and a signal was then piped to the subscribers' houses

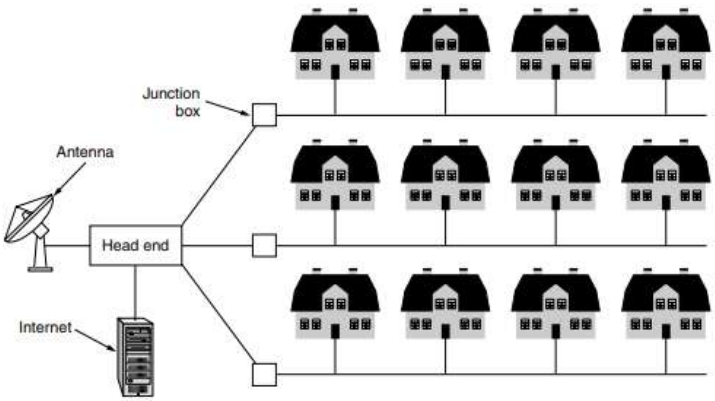


Figure . A metropolitan area network based on cable TV.

4. WAN - Wide Area Networks

- A WAN (Wide Area Network) spans a large geographical area, often a country or continent.
- Wired WANs, using the example of a company with branch offices in different cities.
- The WAN as we have described it looks similar to a large wired LAN, but there are some important differences that go beyond long wires.
- Usually in a WAN, the hosts and subnet are owned and operated by different people.
- In our example, the **employees might be responsible for their own computers**, while the **company's IT department is in charge of the rest of the network**.
- A second difference is that the routers will usually connect different kinds (switched Ethernet, SONET etc.) of networking technology.

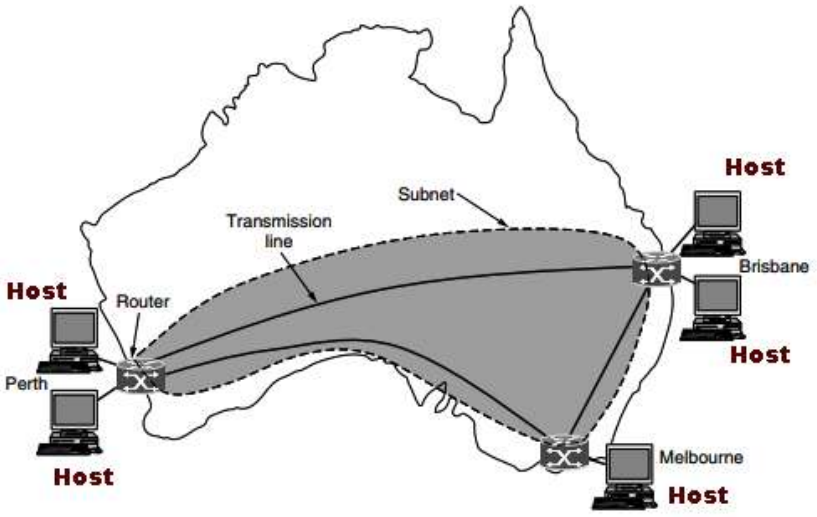


Figure . WAN that connects three branch offices in Australia.

- Some device needs to join them. This means that many WANs will in fact be **internetworks**, or composite networks that are made up of more than one network.
- A final difference is in what is connected to the subnet. This could be individual computers, as was the case for connecting to LANs, or it could be entire LANs. This is how larger networks are built from smaller ones. As far as the subnet is concerned, it does the same job.
- VPN (Virtual Private Network). Compared to the dedicated arrangement, a VPN has the usual advantage of virtualization, which is that it provides flexible reuse of a resource (Internet connectivity). Consider how easy it is to add a fourth office to see this. A VPN also has the usual disadvantage of virtualization, which is a lack of control over the underlying resources. With a dedicated line, the capacity is clear. With a VPN your mileage may vary with your Internet service.

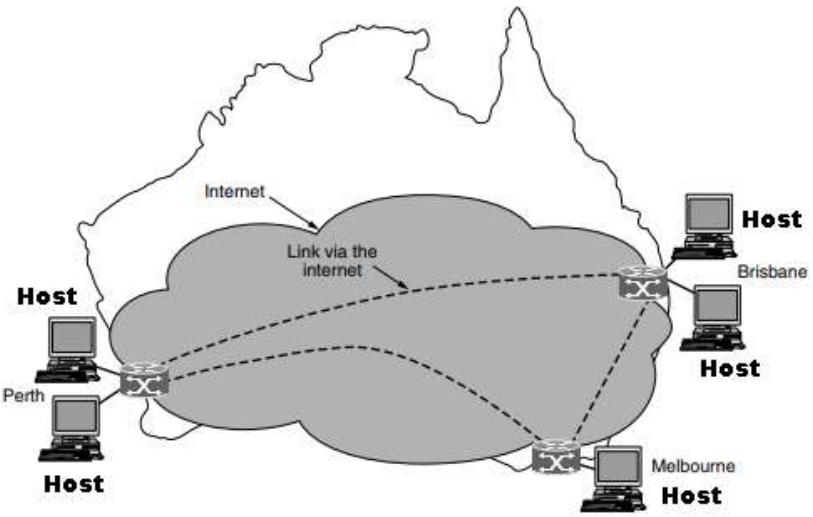


Figure . WAN using a virtual private network.

- The **subnet operator is known as a network service provider** and the offices are its customers. This structure is shown in Fig. The subnet operator will connect to other customers too, as long as they can pay and it can provide service. Since it would be a disappointing network service if the customers could only send packets to each other, the subnet operator will also connect to other networks that

are part of the Internet. Such a subnet operator is called an ISP (Internet Service Provider) and the subnet is an ISP network. Its customers who connect to the ISP receive Internet service.

- In most WANs, the network contains many transmission lines, each connecting a **pair of routers**. If **two routers that do not share a transmission line wish to communicate**, they must do this indirectly, via other routers.
- There may be many paths in the network that connect these two routers. How the network makes the decision as to which path to use is called the **routing algorithm.** Many such algorithms exist. Each router makes the decision as to where to send a packet next is called the **forwarding algorithm**.

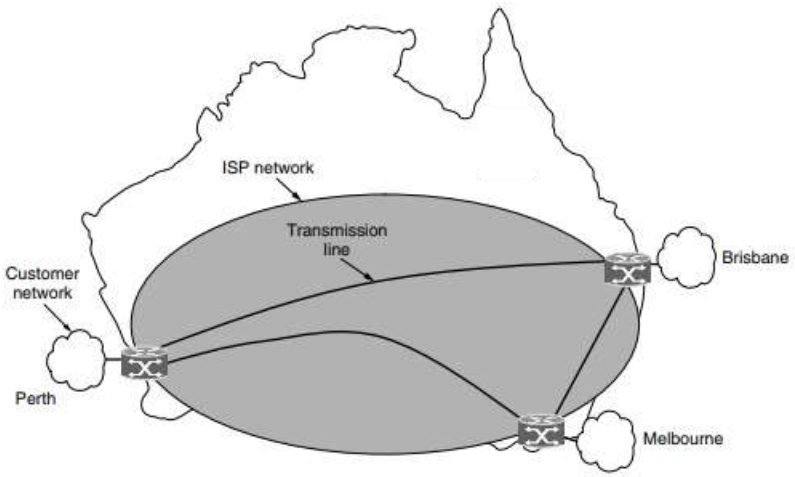


Figure 1-12. WAN using an ISP network.

- Other kinds of **WANs make heavy use of wireless technologies**. In satellite systems, each computer on the ground has an antenna through which it can send data to and receive data from to a satellite in orbit.
- The cellular telephone network is another example of a WAN that uses wireless technology.
    1. The first generation was analog and for voice only.
    2. The second generation was digital and for voice only.
    3. The third generation is digital and is for both voice and data.
- Each cellular base station covers a distance much larger than a wireless LAN with a range measured in kilometers rather than tens of meters.

5. Internetworks or Internet

- People connected to one network often want to communicate with people attached to a different one.
- The fulfilment of this desire requires that different, and frequently incompatible, networks be connected. A collection of interconnected networks is called **an internetwork or internet**.
- The Internet uses ISP networks to connect enterprise networks, home networks, and many other networks.
- A network is formed by the combination of a subnet and its hosts. We know that an internet is formed when distinct networks are interconnected.
- In our view, connecting a LAN and a WAN or connecting two LANs is the usual way to form an internetwork, but there is little agreement in the industry over terminology in this area.
- There are two rules of thumb that are useful.
    - First, if different organizations have paid to construct different parts of the network and each maintains its part, we have an internetwork rather than a single network.
    - Second, if the underlying technology is different in different parts (e.g., broadcast versus point-to-point and wired versus wireless), we probably have an internetwork.
- The general name for a machine that makes a connection between two or more networks and provides the necessary translation, both in terms of hardware and software, is a gateway.
- Gateways are distinguished by the layer at which they operate in the protocol hierarchy.

- Since the benefit of forming an internet is to connect computers across networks, the level in the middle that is ‘‘just right’’ is often called the network layer, and a router is a gateway that switches packets at the network layer. We can now spot an internet by finding a network that has routers.

NETWORK SOFTWARE

Protocol Hierarchies - Design Issues for the Layers - Connection-Oriented Versus Connectionless Service - Service Primitives - The Relationship of Services to Protocols.

Protocol Hierarchies

- To reduce their design complexity, most networks are organized as a stack of layers or levels, each one built upon the one below it.
- The number of layers, the name of each layer, the contents of each layer, and the function of each layer differ from network to network.
- The purpose of each layer is to offer certain services to the higher layers while shielding those layers from the details of how the offered services are actually implemented.
- In a sense, each layer is a kind of virtual machine, offering certain services to the layer above it.
- This concept is variously known as
  - information hiding,
  - abstract data types,
  - data encapsulation, and
  - object-oriented programming.
- The fundamental idea is that a particular piece of software (or hardware) provides a service to its users but keeps the details of its internal state and algorithms hidden from them.
- When layer **“n”** on one machine carries on a conversation with layer **“n”** on another machine, the rules and conventions used in this conversation are collectively known as the **layer n protocol**.
- Basically, a **protocol** is an agreement between the communicating parties on how communication is to proceed.
- A five-layer network is illustrated in Figure. The entities comprising the corresponding layers on different machines are called peers. The peers may be software processes, hardware devices, or even human beings. In other words, it is the peers that communicate by using the protocol to talk to each other.

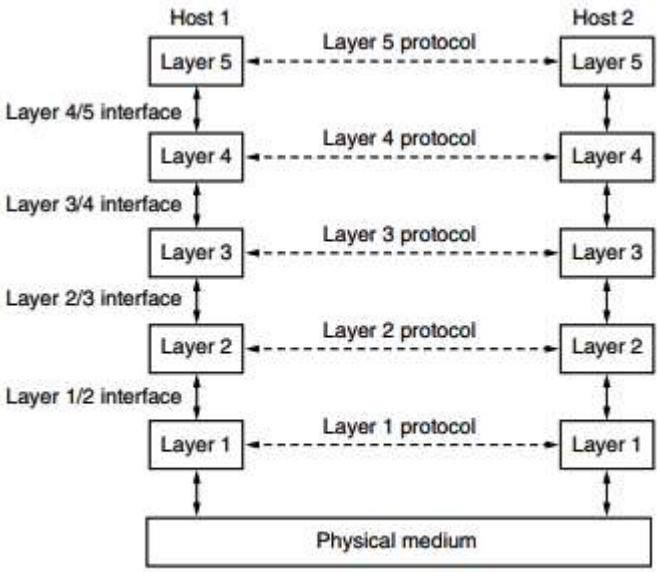


Figure . Layers, protocols, and interfaces.

- Between each pair of adjacent layers is an interface. The interface defines which primitive operations and services the lower layer makes available to the upper one. When network designers decide how many layers to include in a network and what each one should do, one of the most important considerations is defining clean interfaces between the layers.

- A set of layers and protocols is called a **network architecture**. The specification of an architecture must contain enough information to allow an implementer to write the program or build the hardware for each layer so that it will correctly obey the appropriate protocol.
- A list of the protocols used by a certain system, one protocol per layer, is called a **protocol stack**.
- Now consider a more technical example: how to provide communication to the top layer of the five-layer network shown in Figure. A message, M, is produced by an application process running in layer 5 and given to layer 4 for transmission. Layer 4 puts a header in front of the message to identify the message and passes the result to layer 3. The header includes control information, such as addresses, to allow layer 4 on the destination machine to deliver the message. Other examples of control information used in some layers are sequence numbers (in case the lower layer does not preserve message order), sizes, and times.
- The important thing to understand about the Figure is the relation between the virtual and actual communication and the difference between protocols and interfaces. The peer processes in layer 4, for example, conceptually think of their communication as being ''horizontal,'' using the layer 4 protocol. Each one is likely to have procedures called something like **SendToOtherSide** and **GetFromOtherSide**, even though these procedures actually communicate with lower layers across the 3/4 interface, and not with the other side.

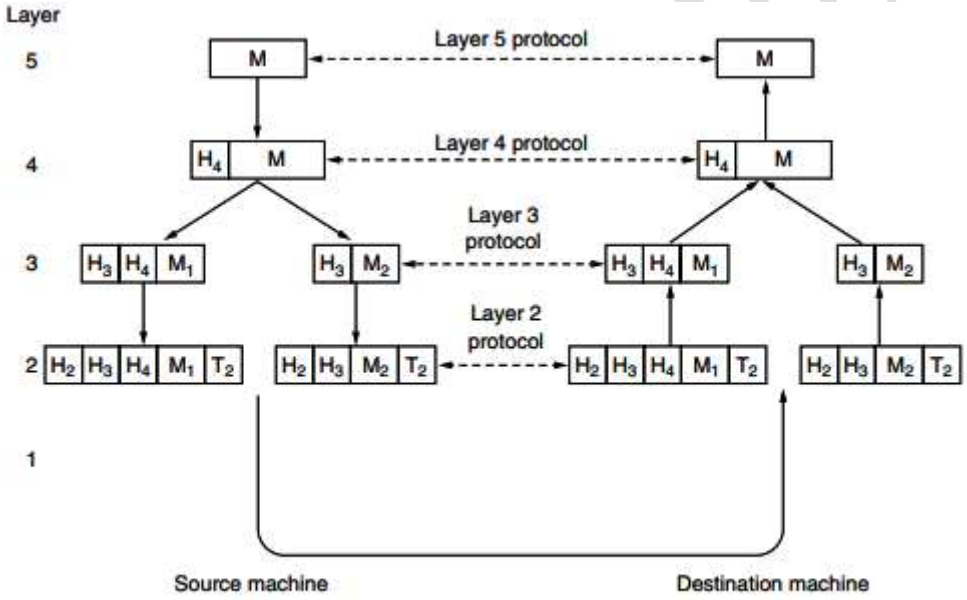


Figure. Example information flow supporting virtual communication in layer 5.

*****

Design Issues for the Layers

1. The bits of a packet traveling through the network. will be received damaged (inverted) due to fluke electrical noise, random wireless signals, hardware flaws, software bugs and so on. One mechanism for finding errors in received information uses codes for **error detection** and more powerful codes allow for **error correction.**
2. Reliability issue is finding a working path through a network - The network should automatically make this decision to correct the broken routes - called **routing**.
3. Design issue concerns the evolution of the network - Over time, networks grow larger and new designs emerge that need to be connected to the existing network. We have recently seen the key

structuring mechanism used to support change by dividing the overall problem and hiding implementation details - **protocol layering**.

4. Next design issue is **resource allocation**. Networks provide a service to hosts from their underlying resources, such as the capacity of transmission lines. To do this well, they need mechanisms that divide their resources so that one host does not interfere with another too much.
5. The last major design issue is to **secure the network by defending it against different kinds of threats.** Mechanisms that provide confidentiality defend against this threat, and they are used in multiple layers. Mechanisms for authentication prevent someone from impersonating someone else.

*****

Connection-Oriented Versus Connectionless Service

- Layers can offer two different types of service to the layers above them:
    1. connection-oriented service and
    2. Connectionless service.

Connection-oriented service

- Connection-oriented service is modelled after the telephone system.
- To use a connection-oriented network service,
    - the service user first establishes a connection
    - uses the connection, and then
    - releases the connection.
- The essential aspect of a connection is that it acts like a tube: the sender pushes objects (bits) in at one end, and the receiver takes them out at the other end. In most cases the order is preserved so that the bits arrive in the order they were sent.
- In some cases when a connection is established, the sender, receiver, and subnet conduct a negotiation about the parameters to be used, such as maximum message size, quality of service required, and other issues.
- Typically, one side makes a proposal and the other side can accept it, reject it, or make a counter proposal.
- A **circuit** is another name for a connection with associated resources, such as a fixed bandwidth. This dates from the telephone network in which a circuit was a path over copper wire that carried a phone conversation.

Connectionless service

- Connectionless service is modelled after the postal system. Each message (letter) carries the full destination address, and each one is routed through the intermediate nodes inside the system independent of all the subsequent messages.
- There are different names for messages in different contexts; a packet is a message at the network layer.
- When the intermediate nodes receive a message in full before sending it on to the next node, this is called **store-and-forward switching**.
- The alternative, in which the onward transmission of a message at a node starts before it is completely received by the node, is called **cut-through switching**.
- Each kind of service can further be characterized by its reliability. Some services are reliable in the sense that they never lose data.

- Usually, a reliable service is implemented by having the receiver **acknowledge the receipt** of each message so the sender is sure that it arrived. The **acknowledgement process** introduces overhead and delays, which are often worth it but are sometimes undesirable.
- A typical situation in which a reliable connection-oriented service is appropriate is file transfer.
- Reliable connection-oriented service has **two minor variations**: **message sequences** and **byte streams.**
- The following table summarizes the types of services:

	Service	Example
Connection-oriented	Reliable message stream	Sequence of pages
	Reliable byte stream	Movie download
	Unreliable connection	Voice over IP
Connection-less	Unreliable datagram	Electronic junk mail
	Acknowledged datagram	Text messaging
	Request-reply	Database query

Figure . Six different types of service.

*****

Service Primitives

- A service is formally specified by a set of primitives (operations) available to user processes to access the service. These primitives tell the service to perform some action or report on an action taken by a peer entity.
- If the protocol stack is located in the operating system, as it often is, the primitives are normally system calls. These calls cause a trap to kernel mode, which then turns control of the machine over to the operating system to send the necessary packets.
- The set of primitives available depends on the nature of the service being provided. The primitives for connection-oriented service are different from those of connectionless service. As a minimal example of the service primitives that might provide a reliable byte stream.

Primitive	Meaning
LISTEN	Block waiting for an incoming connection
CONNECT	Establish a connection with a waiting peer
ACCEPT	Accept an incoming connection from a peer
RECEIVE	Block waiting for an incoming message
SEND	Send a message to the peer
DISCONNECT	Terminate a connection

Figure . Six service primitives that provide a simple connection-oriented service.

- These primitives might be used for a request-reply interaction in a client-server environment.

- To illustrate how, We sketch a simple protocol that implements the service using acknowledged datagrams:

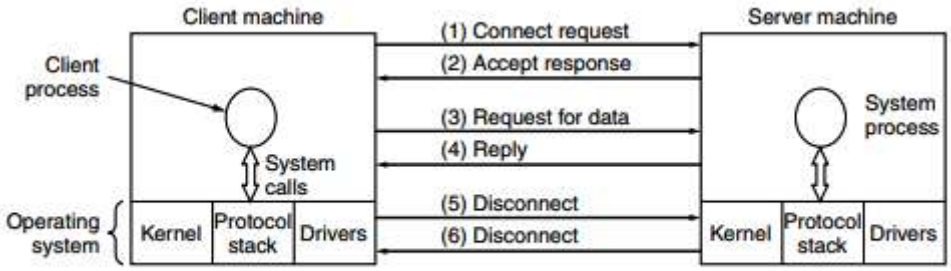


Figure. A simple client-server interaction using acknowledged datagrams.

*****

The Relationship of Services to Protocols.

- Services and protocols are distinct concepts. The service and the protocol are completely decoupled.
- Services relate to the interfaces between layers, In contrast, protocols relate to the packets sent between peer entities on different machines.
- A **service**
  - is a set of primitives (operations) that a layer provides to the layer above it.
  - defines what operations (not how) the layer is prepared to perform on behalf of its users.
  - relates to an interface between two layers, with the **lower layer being the service provider** and the **upper layer being the service user**.
- A protocol
  - is a set of rules governing the format and meaning of the packets, or messages that are exchanged by the peer entities within a layer.
  - Entities use protocols to implement their service definitions.
  - They are free to change their protocols at will, provided they do not change the service visible to their users.

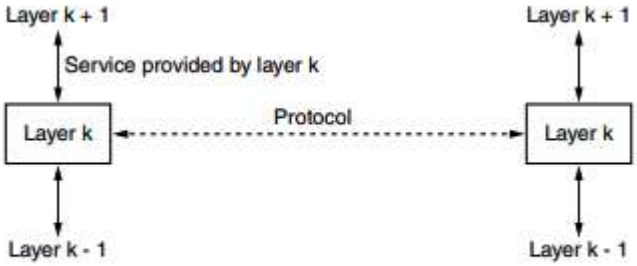


Figure. The relationship between a service and a protocol.

REFERENCE MODELS

- **The OSI Reference Model - The TCP/IP Reference Model**
- **A Comparison of the OSI and TCP/IP Reference Models**
- **A Critique of the OSI Model and Protocols - A Critique of the TCP/IP Reference Model.**

The OSI Reference Model

- OSI (Open Systems Interconnection) Reference Model because it deals with connecting open systems—that is, systems that are open for communication with other systems.

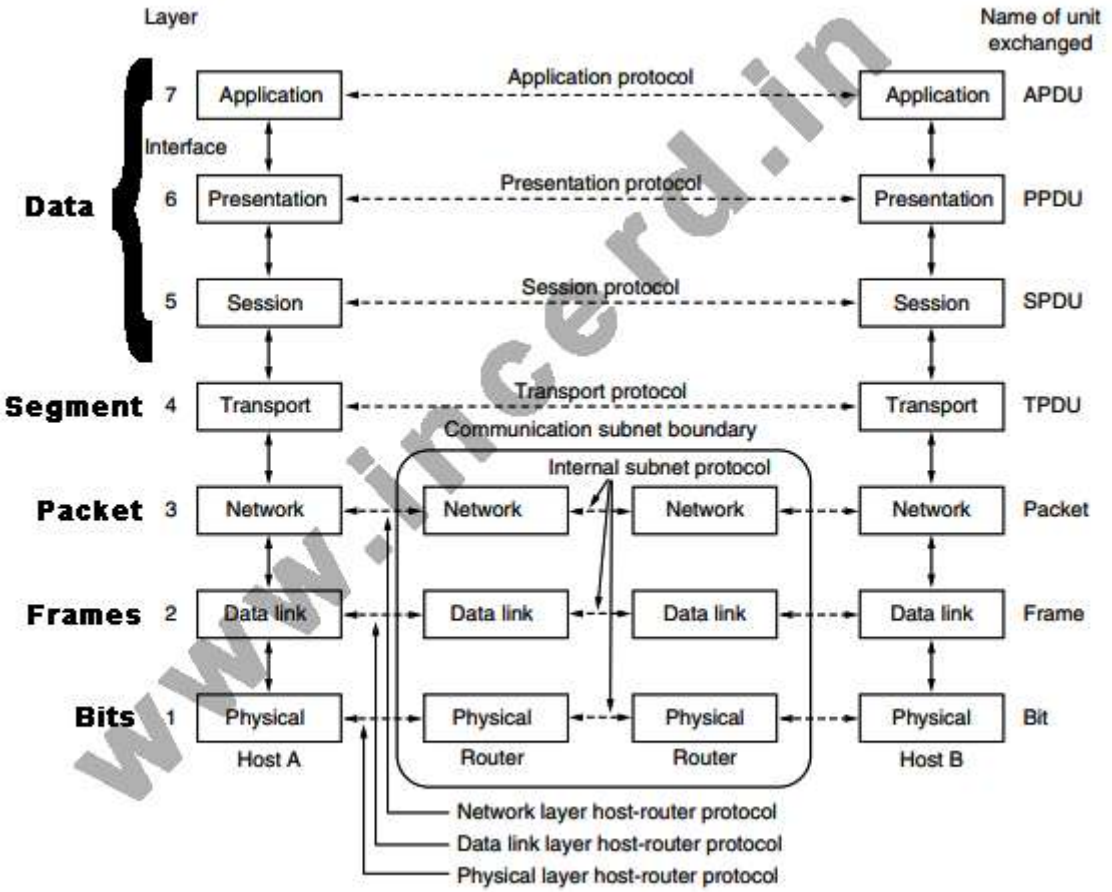
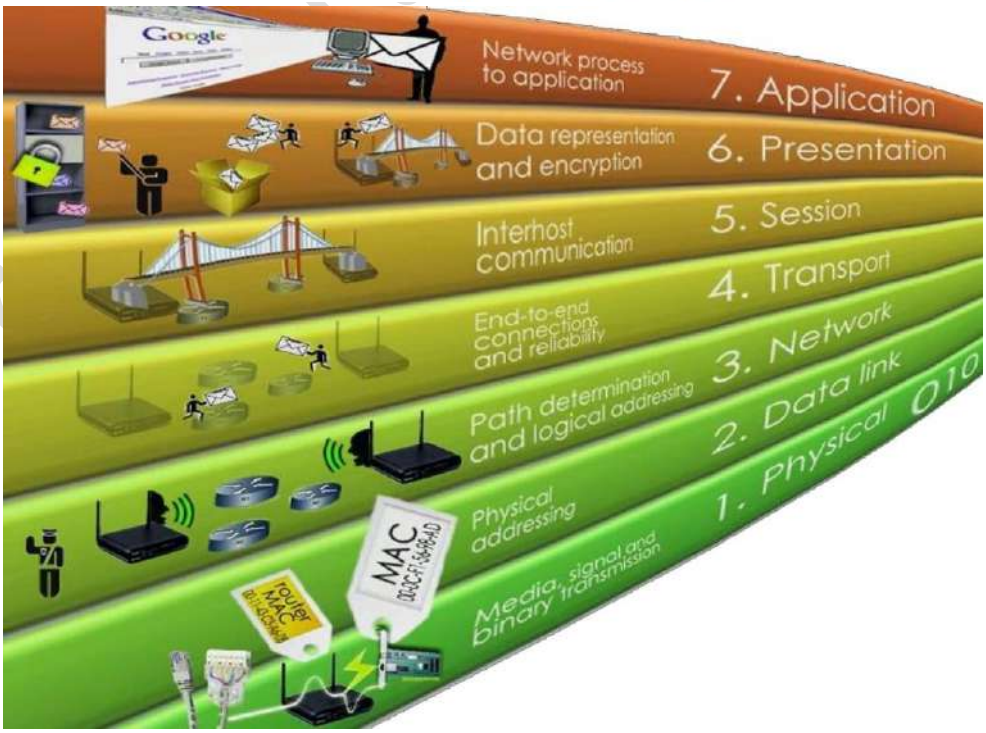


Figure . The OSI reference model.



- The OSI model has seven layers. The principles that were applied to arrive at the seven layers can be briefly summarized as follows:
    - A layer should be created where a different abstraction is needed.
    - Each layer should perform a well-defined function.
    - The function of each layer should be chosen with an eye toward defining internationally standardized protocols.

1. The **physical layer** is concerned with transmitting raw bits over a communication channel. The design issues have to do with making sure that when one side sends a 1 bit it is received by the other side as a 1 bit, not as a 0 bit.
2. The main task of the **data link layer** is to transform a raw transmission facility into a line that appears free of undetected transmission errors. It does so by masking the real errors so the network layer does not see them. It accomplishes this task by having the sender break up the input data into data frames (typically a few hundred or a few thousand bytes) and transmit the frames sequentially. If the service is reliable, the receiver confirms correct receipt of each frame by sending back an acknowledgement frame. Broadcast networks have an additional issue in the data link layer: how to control access to the shared channel. A special sublayer of the data link layer, the **medium access control sublayer**, deals with this problem.
3. The **network layer** controls the operation of the subnet. A key design issue is determining how packets are routed from source to destination. Routes can be based on static tables that are ''wired into'' the network and rarely changed, or more often they can be updated automatically to avoid failed components. They can also be determined at the start of each conversation, for example, a terminal session, such as a login to a remote machine. Finally, they can be highly dynamic, being determined anew for each packet to reflect the current network load.
4. The basic function of the **transport layer** is to accept data from above it, split it up into smaller units if need be, pass these to the network layer, and ensure that the pieces all arrive correctly at the other end. Furthermore, all this must be done efficiently and in a way that isolates the upper layers from the inevitable changes in the hardware technology over the course of time.
5. The **session layer** allows users on different machines to establish sessions between them. Sessions offer various services, including dialog control (keeping track of whose turn it is to transmit), token management (preventing two parties from attempting the same critical operation simultaneously), and synchronization (check pointing long transmissions to allow them to pick up from where they left off in the event of a crash and subsequent recovery).
6. The **presentation layer** is concerned with the syntax and semantics of the information transmitted. In order to make it possible for computers with different internal data representations to communicate, the data structures to be exchanged can be defined in an abstract way, along with a standard encoding to be used ''on the wire.'' The presentation layer manages these abstract data structures and allows higher-level data structures (e.g., banking records) to be defined and exchanged.
7. The **application layer** contains a variety of protocols that are commonly needed by users. One widely used application protocol is HTTP (HyperText Transfer Protocol), which is the basis for the World Wide Web. When a browser wants a Web page, it sends the name of the page it wants to the server hosting the page using HTTP. The server then sends the page back. Other application protocols are used for file transfer, electronic mail, and network news.

*****

The TCP/IP Reference Model

- Let us now turn from the OSI reference model to the reference model used in the grandparent of all wide area computer networks, the ARPANET, and its successor, the worldwide Internet.
- The ARPANET was a research network sponsored by the DoD (U.S. Department of Defense). It eventually connected hundreds of universities and government installations, using leased telephone lines.
- When satellite and radio networks were added later, the existing protocols had trouble interworking with them, so a new reference architecture was needed.
- Thus, from nearly the beginning, the ability to connect multiple networks in a seamless way was one of the major design goals. This architecture later became known as the TCP/IP Reference Model, after its two primary protocols.
- It was first described by Cerf and Kahn (1974), and later refined and defined as a standard in the Internet community (Braden, 1989). The design philosophy behind the model is discussed by Clark (1988)

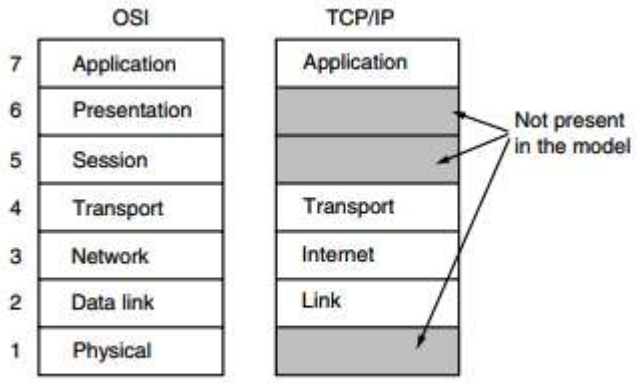


Figure . The TCP/IP reference model.

1. **The Link Layer**: All these requirements led to the choice of a packet-switching network based on a connectionless layer that runs across different networks. The lowest layer in the model, **the link layer describes what links such as serial lines and classic Ethernet** must do to meet the needs of this connectionless internet layer.
2. **The Internet Layer**
   - The internet layer is the hub that holds the whole architecture together. It is shown in the above figure as corresponding roughly to the OSI network layer. Its job is to permit hosts to **inject packets into any network** and have them **travel independently to the destination** (potentially on a different network). They may **even arrive in a completely different order** than they were sent, in which case it is the job of **higher layers to rearrange them**, if in-order delivery is desired.
   - The internet layer defines an official packet format and **protocol called IP (Internet Protocol),** plus a companion protocol called **ICMP (Internet Control Message Protocol)** that helps it function. The job of the internet layer is to deliver IP packets where they are supposed to go. **Packet routing is clearly a major issue here, as is congestion** (though IP has not proven effective at avoiding congestion).
3. **The Transport Layer**
   - The layer above the internet layer in the TCP/IP model is now usually called the transport layer. It is designed to allow peer entities on the source and destination hosts to carry on a

conversation, just as in the OSI transport layer. Two end-to-end transport protocols have been defined here.

- **The first one, TCP (Transmission Control Protocol)**, is a **reliable connection-oriented protocol** that allows a **byte stream** originating on one machine to be delivered without error on any other machine in the internet. It segments the incoming byte stream into discrete messages and passes each one on to the internet layer. At the destination, **the receiving TCP process reassembles the received messages into the output stream**. **TCP also handles flow control** to make sure a fast sender cannot swamp a slow receiver with more messages than it can handle.
- The second protocol in this layer, **UDP (User Datagram Protocol), is an unreliable, connectionless protocol** for applications that do not want TCP's sequencing or flow control and wish to provide their own. It is also widely used for one-shot, client-server-type request-reply queries and applications in which prompt delivery is **more important** than accurate delivery, such as **transmitting speech or video**. The relation of IP, TCP, and UDP is shown in the following figure. Since the model was developed, IP has been implemented on many other networks.

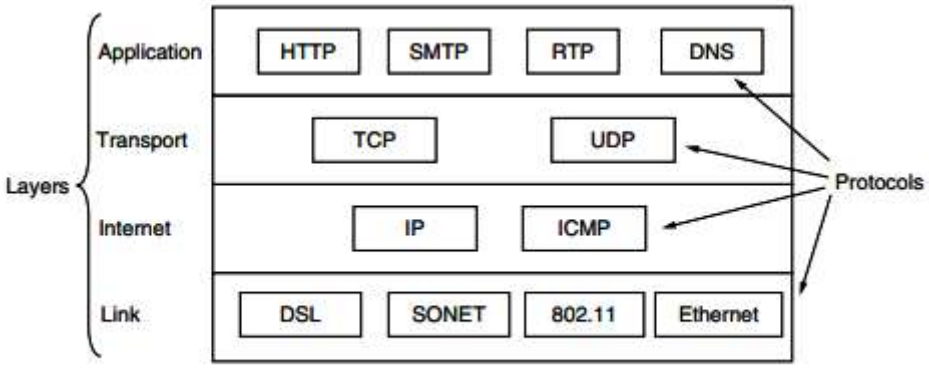


Figure. The TCP/IP model with some protocols we will study.

4. **The Application Layer:** The TCP/IP model does not have session or presentation layers. It contains all the higher-level protocols. The early ones included **virtual terminal (TELNET), file transfer (FTP),** and **electronic mail (SMTP)**. Some important ones the **Domain Name System (DNS)**, for mapping host names onto their network addresses, **HTTP (Hyper Text Transfer Protocol) , the protocol** for fetching pages on the World Wide Web, and **RTP (Real-time Transport Protocol)**, the protocol for delivering real-time media such as voice or movies.

*****

A Comparison of the OSI and TCP/IP Reference Models

- Three concepts are central to **the OSI model**:
  1. Services
  2. Interfaces
  3. Protocols
- Each layer performs some **services** for the layer above it. The service definition tells what the layer does, not how entities above it access it or how the layer works. It defines the layer's semantics.
- A layer's **interface** tells the processes above it how to access it. It specifies what the parameters are and what results to expect.

- Finally, the peer **protocols** used in a layer are the layer's own business. It can use any protocols it wants to, as long as it gets the job done (i.e., provides the offered services). It can also change them at will without affecting software in higher layers.
- The OSI reference model was devised before the corresponding protocols were invented.
- **The TCP/IP model** did not originally clearly distinguish between services, interfaces, and protocols.
- With TCP/IP, the protocols came first, and the model was really just a description of the existing protocols.
- The number of layers: the OSI model has seven layers and the TCP/IP model has four.
- Another difference is in the area of connectionless versus connection-oriented communication. The OSI model supports both connectionless and connection-oriented communication in the network layer, but only connection-oriented communication in the transport layer, where it counts (because the transport service is visible to the users).

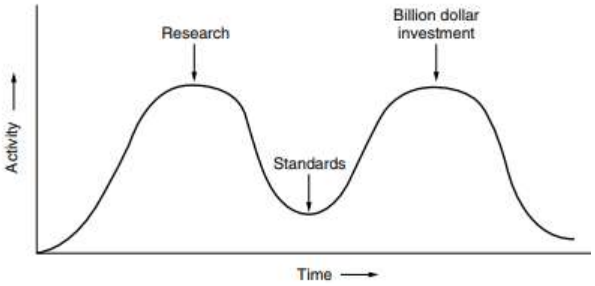
*****

A Critique of the OSI Model and Protocols

1. Bad timing
2. Bad technology
3. Bad implementations
4. Bad politics

1. Bad timing

- First let us look at reason one: bad timing. **The time** at which a standard is established is absolutely **critical to its success**.
- This figure shows the amount of activity surrounding a new subject. When the subject is first discovered, there is a burst of research activity in the form of discussions, papers, and meetings.



- After a while this activity subsides, corporations discover the subject, and the billion-dollar wave of investment hits.
- It now appears that the standard OSI protocols got crushed. The competing TCP/IP protocols were already in widespread use by research universities by the time the OSI protocols appeared.

2. Bad technology

- The second reason that OSI never caught on is that both the model and the protocols are flawed.
- The choice of seven layers was more political than technical, and two of the layers (session and presentation) are nearly empty, whereas two other ones (data link and network) are overfull.
- The OSI model, along with its associated service definitions and protocols, is extraordinarily complex. When piled up, the printed standards occupy a significant fraction of a meter of paper.
- They are also difficult to implement and inefficient in operation.

3. Bad implementations

- Given the enormous complexity of the model and the protocols, it will come as no surprise that the initial implementations were huge, unwieldy, and slow.
- Everyone who tried them got burned. It did not take long for people to associate ‘‘OSI’’ with ‘‘poor quality.’’
- Although the products improved in the course of time, the image stuck. In contrast, one of the first implementations of TCP/IP was part of Berkeley UNIX and was quite good (not to mention, free).
- People began using it quickly, which led to a large user community, which led to improvements, which led to an even larger community. Here the spiral was upward instead of downward.

4. Bad politics

- On account of the initial implementation, many people, especially in academia, thought of TCP/IP as part of UNIX, and UNIX in the 1980s in academia was not unlike parenthood (then incorrectly called motherhood) and apple pie.
- OSI, on the other hand, was widely thought to be the creature of the European telecommunication ministries, the European Community, and later the U.S. Government.
- This belief was only partly true, but the very idea of a bunch of government bureaucrats trying to shove a technically inferior standard down the throats of the poor researchers and programmers down in the trenches actually developing computer networks did not aid OSI’s cause.

*****

A Critique of the TCP/IP Reference Model.

- The TCP/IP model and protocols have their problems too.
- **First**, the model does not clearly distinguish the concepts of services, interfaces, and protocols.
- **Second**, the TCP/IP model is not at all general and is poorly suited to describing any protocol stack other than TCP/IP. Trying to use the TCP/IP model to describe Bluetooth, for example, is completely impossible.
- **Third**, the link layer is not really a layer at all in the normal sense of the term as used in the context of layered protocols. It is an interface (between the network and data link layers). The distinction between an interface and a layer is crucial, and one should not be disordered about it.
- **Fourth**, the TCP/IP model does not distinguish between the physical and data link layers. These are completely different. The physical layer has to do with the transmission characteristics of copper wire, fiber optics, and wireless communication. The data link layer’s job is to delimit the start and end of frames and get them from one side to the other with the desired degree of reliability. A proper model should include both as separate layers.
- **Finally**, although the IP and TCP protocols were carefully thought out and well implemented, many of the other protocols were ad hoc, generally produced by a couple of graduate students hacking away until they got tired.

- Media are roughly grouped into
  - Guided media, such as copper wire and fiber optics, and
  - Unguided media, such as terrestrial wireless, satellite, and lasers through the air.

Guided transmission media

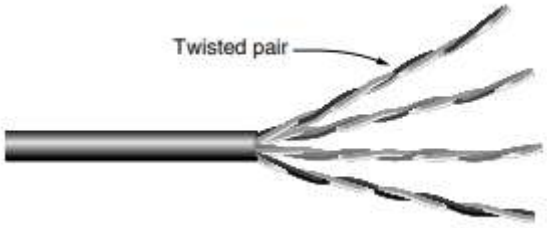
1. Magnetic Media
2. Twisted Pairs
3. Coaxial Cable
4. Power Lines
5. Fiber Optics

1. Magnetic Media

- One of the most common ways to transport data from one computer to another is to write them onto magnetic tape or removable media (e.g., recordable DVDs), physically transport the tape or disks to the destination machine, and read them back in again.

2. Twisted Pairs

- Although the bandwidth characteristics of magnetic tape are excellent, the delay characteristics are poor.
- Transmission time is measured in minutes or hours, not milliseconds. For many applications an online connection is needed.
- One of the oldest and still most common transmission media is twisted pair.
- A twisted pair consists of two insulated copper wires, typically about 1 mm thick. The wires are twisted together in a helical form, just like a DNA molecule.
- Twisting is done because two parallel wires constitute a fine antenna. When the wires are twisted, the waves from different twists cancel out, so the wire radiates less effectively.
- A signal is usually carried as the difference in voltage between the two wires in the pair.
- This provides better immunity to external noise because the noise tends to affect both wires the same, leaving the differential unchanged.

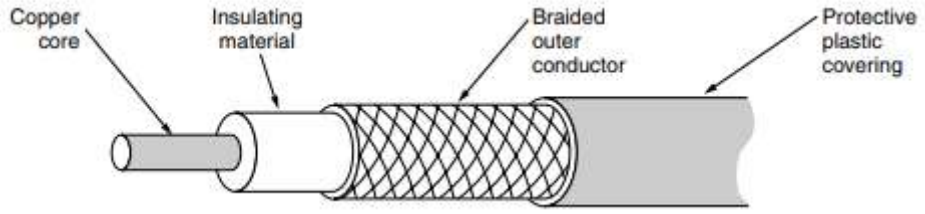


- The most common application of the twisted pair is the telephone system.
- Both telephone calls and ADSL Internet access run over these lines. Twisted pairs can run several kilometers without amplification, but for longer distances the signal becomes too attenuated and repeaters are needed.
- Links that can be used in **both directions at the same time**, like a two-lane road, are called **full-duplex links**.
- In contrast, links that can be **used in either direction**, but only one way at a time, like a single-track railroad line are called **half-duplex links**.
- A third category consists of links that **allow traffic in only one direction**, like a one-way street. They are called **simplex links**.
- Returning to twisted pair, **Cat 5** replaced earlier **Category 3** cables with a similar cable that uses the same connector, but has more twists per meter. More twists result in less crosstalk and a better-quality signal over longer distances, making the cables more suitable for high-speed computer communication, especially 100-Mbps and 1-Gbps Ethernet LANs.
- New wiring is more likely to be **Category 6** or even **Category 7**. These categories has more stringent specifications to handle signals with greater bandwidths.

- Some cables in Category 6 and above are rated for signals of 500 MHz and can support the 10-Gbps links that will soon be deployed. Through **Category 6, these wiring types are referred to as UTP (Unshielded Twisted Pair)** as they consist simply of wires and insulators. In contrast to these, **Category 7 cables have shielding on the individual twisted pairs, as well as around the entire cable (but inside the plastic protective sheath).**

3. Coaxial Cable

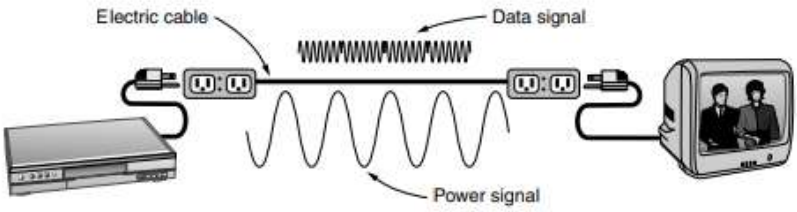
- Another common transmission medium is the coaxial cable (known to its many friends as just ''coax'' and pronounced ''co-ax''). It has better shielding and greater bandwidth than unshielded twisted pairs, so it can span longer distances at higher speeds.
- Two kinds of coaxial cable are widely used. One kind, 50-ohm cable, is commonly used when it is intended for digital transmission from the start. The other kind, 75-ohm cable, is commonly used for analog transmission and cable television. This **distinction is based on historical, rather than technical, factors** (e.g., early dipole antennas had an impedance of 300 ohms, and it was easy to use existing 4:1 impedance-matching transformers). Starting in the mid1990s, cable TV operators began to provide Internet access over cable, which has made 75-ohm cable more important for data communication.
- A coaxial cable consists of a stiff copper wire as the core, surrounded by an insulating material. The insulator is encased by a cylindrical conductor, often as a closely woven braided mesh. The outer conductor is covered in a protective plastic sheath. A cutaway view of a coaxial cable is shown in Figure.



- The construction and shielding of the coaxial cable give it a good combination of high bandwidth and excellent noise immunity.

4. Power Lines

- The telephone and cable television networks are not the only sources of wiring that can be reused for data communication. There is a yet more common kind of wiring: **electrical power lines**. Power lines deliver electrical power to houses, and electrical wiring within houses distributes the power to electrical outlets.



- The convenience of using power lines for networking should be clear. Simply plug a TV and a receiver into the wall, which you must do anyway because they need power, and they can send and receive movies over the electrical wiring.
- This configuration is shown in Figure below. There is no other plug or radio. The data signal is superimposed on the low-frequency power signal (on the active or ''hot'' wire) as both signals use the wiring at the same time.

- It is practical to send at least 100 Mbps over typical household electrical wiring by using communication schemes that resist impaired frequencies and bursts of errors. Many products use various proprietary standards for power-line networking, so international standards are actively under development

5. Fiber Optics

- Fiber optics are used for long-haul transmission in network backbones, highspeed LANs (although so far, copper has always managed catch up eventually), and high-speed Internet access such as **FttH (Fiber to the Home).**
- An optical transmission system has **three key components**:
  I. the light source
  II. the transmission medium and
  III. the detector.
- Conventionally, a **pulse of light indicates a 1 bit** and the **absence of light indicates a 0 bit**.
- The transmission medium is an **ultra-thin fiber of glass**. The detector generates an electrical pulse when light falls on it. By attaching a light source to one end of an optical fiber and a detector to the other, we have a unidirectional data transmission system that accepts an electrical signal, converts and transmits it by light pulses, and then reconverts the output to an electrical signal at the receiving end.

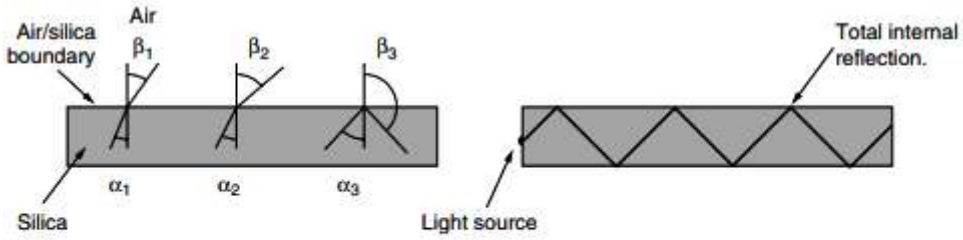


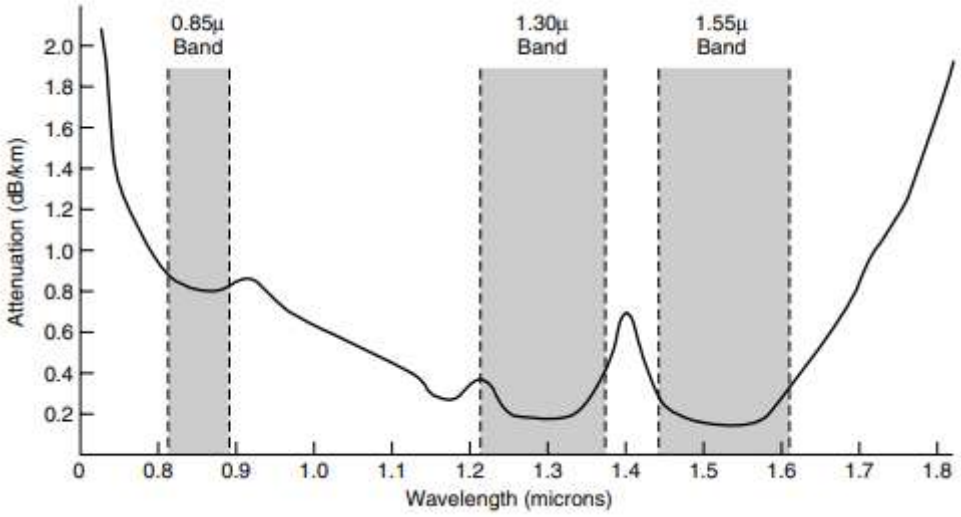
Figure . (a)
Three examples of a light ray from inside a silica fiber imping ing on the air/silica boundary at different angles.

Figure . (b)
Light trapped by total inter nal reflection.

- The sketch of Fig. 6(b) shows only one trapped ray, but since any light ray incident on the boundary above the critical angle will be reflected internally, many different rays will be bouncing around at different angles. Each ray is said to have a different mode, so a fiber having this property is called a **multimode fiber**.
- However, if the **fiber's diameter is reduced to a few wavelengths of light** the fiber acts like a **wave guide and the light can propagate only in a straight line**, **without bouncing, yielding** a **single-mode fiber**.
- **Single-mode fibers are more expensive** but are widely **used for longer distances**. Currently available single-mode fibers can transmit data at 100 Gbps for 100 km without amplification. Even higher data rates have been achieved in the laboratory for shorter distances.

Transmission of Light through Fiber

- Optical fibers are made of glass, which, in turn, is made from sand, an inexpensive raw material available in unlimited amount.
- The attenuation of light through glass depends on the wavelength of the light (as well as on some physical properties of the glass). It is defined as the ratio of input to output signal power.
- For the kind of glass used in fibers, the attenuation is shown in Figure below, in units of decibels per linear kilometer of fiber.



- **Three wavelength bands** are most commonly used at present for optical communication. They are centered at **0.85, 1.30, and 1.55 microns**, respectively. **All three bands are 25,000 to 30,000 GHz wide.**
- The **0.85-micron band** was used first. It has higher attenuation and so is **used for shorter distances**, but at that wavelength the lasers and electronics could be made from the same material (gallium arsenide).
- The last two bands have good attenuation properties (less than 5% loss per kilometer). The 1.55-micron band is now widely used with erbium-doped amplifiers that work directly in the optical domain.
- **Light pulses sent down a fiber spread out in length as they propagate**. This spreading is called **chromatic dispersion**. The amount of it is wavelength dependent.
- These pulses are called **solitons**.

Fiber Cable

- Fiber optic cables are similar to coax, except without the braid.
- Figure (a) shows a single fiber viewed from the side. At the center is the glass core through which the light propagates. In **multimode fibers**, **the core is typically 50 microns in diameter**, about the thickness of a human hair. In **single-mode fibers**, the core is **8 to 10 microns.**
- The core is **surrounded by a glass cladding** with a lower index of refraction than the core, to keep all the light in the core. Next comes a thin plastic jacket to protect the cladding.
- Fibers are typically grouped in bundles, protected by an outer sheath. Figure (b) shows a sheath with three fibers.
- Terrestrial fiber sheaths are normally laid in the ground within a meter of the surface, where they are occasionally subject to attacks by backhoes or gophers. Near the shore, transoceanic fiber sheaths are buried in trenches by a kind of seaplow. In deep water, they just lie on the bottom, where they can be snagged by fishing trawlers or attacked by giant squid.

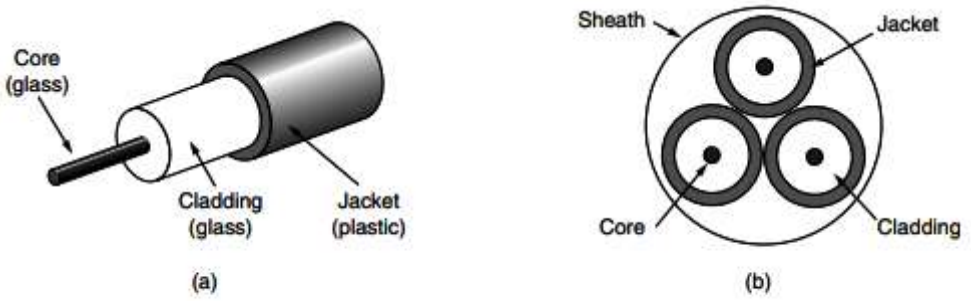


Figure . (a) Side view of a single fiber. (b) End view of a sheath with three fibers.

- Fibers can be connected in three different ways.
  I. First, they can terminate in connectors and be plugged into fiber sockets. Connectors lose about 10 to 20% of the light, but they make it easy to reconfigure systems.
  II. Second, they can be spliced mechanically. Mechanical splices just lay the two carefully cut ends next to each other in a special sleeve and clamp them in place. Alignment can be improved by passing light through the junction and then making small adjustments to maximize the signal. Mechanical splices take trained personnel about 5 minutes and result in a 10% light loss.
  III. Third, two pieces of fiber can be fused (melted) to form a solid connection. A fusion splice is almost as good as a single drawn fiber, but even here, a small amount of attenuation occurs.
- For all three kinds of splices, reflections can occur at the point of the splice, and the reflected energy can interfere with the signal.
- **Two kinds of light sources** are typically used to do the signalling. These are
  I. LEDs (Light Emitting Diodes) and
  II. Semiconductor lasers.

Item	LED	Semiconductor laser
Data rate	Low	High
Fiber type	Multi-mode	Multi-mode or single-mode
Distance	Short	Long
Lifetime	Long life	Short life
Temperature sensitivity	Minor	Substantial
Cost	Low cost	Expensive

- They can be tuned in wavelength by inserting **Fabry-Perot** or **Mach-Zehnder interferometers** between the source and the fiber.
- **Fabry-Perot interferometers** are **simple resonant cavities consisting of two parallel mirrors**. The light is incident perpendicular to the mirrors. The length of the cavity selects out those wavelengths that fit inside an integral number of times.
- **Mach-Zehnder interferometers** separate the **light into two beams**. The two beams travel slightly different distances. They are recombined at the end and are in phase for only certain wavelengths.
- The **receiving end** of an optical fiber consists of a **photodiode**, **which gives off an electrical pulse when struck by light**. The response time of photodiodes, which convert the signal from the optical to the electrical domain, limits **data rates to about 100 Gbps**.
- **Thermal noise is also an issue**, so a pulse of light must carry enough energy to be detected. **By making the pulses powerful enough, the error rate can be made arbitrarily small**.

Comparison of Fiber Optics and Copper Wire

Sr. No.	Basis	Fiber Optic Cable	Copper Wire
1.	Data Carrier	It carries data in the form of light.	It carries data in the form of electric signals.
2.	Bandwidth	It offers higher bandwidth.	It offers lower bandwidth.
3.	Structure	It is thin, lighter in weight, and smaller in size.	It is heavier and thicker.
4.	Environment	It can be laid in different environments because it is more resistant to corrosive materials.	It cannot be laid in a different environment because it is more prone to corrosive materials.
5.	Attenuation	Attenuation is very low.	Attenuation is high.
6.	Interface	As this data travel in the form of light, they are not affected by the electrical and magnetic interfaces.	As in this data travel in the form of electric signals, they are affected by the electrical and magnetic interfaces.
7.	Security	They provide security against the wiretappers, because there is no leakage of light and are difficult to tap.	They do not provide security against the wiretappers, because there is leakage of signals, and are easy to tap.
8.	Cross-talk problem	There is no such kind of problem.	These are prevalent this problem.
9.	Effect on charge carriers	In this charge carriers are photons, they do not carry any charge, so they do not get affected.	In this charge carriers are electrons, they carry a negative charge, so they get affected when they move in a wire.
10.	Break-ability	They are easily breakable.	They cannot be easily broken.
11.	Installation Cost	Installation Cost is high.	Installation Cost is less.
12.	Bandwidth Size	It is a bandwidth size 60Tps.	It is a bandwidth size 10Gbps.
13.	Width	Fiber Optic width around 4lbs/1000 ft.	Copper wire width around 39lbs/1000ft.

Attenuation in networking

- Attenuation in computer networking is the loss of communication signal strength that is measured in decibels. As the rate of attenuation increases, the transmission, such as a phone call or an email a user tries to send, becomes more distorted.
- Attenuation occurs on computer networks because of the following factors:
  I. **Range**. Both wired and wireless transmissions gradually dissipate in strength over longer distances.
  II. **Interference**. Radio interference or physical obstructions, such as walls, dampen communication signals on wireless networks.
  III. **Wire size**. Thinner wires suffer from more attenuation than thicker wires on wired networks.

***

Digital Modulation and Multiplexing

Introduction

Digital modulation is the process of converting between digital bits and analog signals (such as a continuously varying voltage)

Schemes that directly encode bits into signals result in **baseband transmission**, where the signal occupies frequencies from zero up to a maximum that depends on the signaling rate.

Passband transmission is where signals are shifted so that "the signal occupies a band of frequencies around the frequency of the carrier signal."

Channels are often shared by multiple signals. Sharing signals is called **multiplexing**. There are several ways to perform multiplexing, including time, frequency, and code division multiplexing.

Baseband transmission

Baseband transmission is the transmission of a raw signal that occupies frequencies from zero up to a maximum that depends on the signaling rate.

NRZ (Non-Return-to-Zero) is a simple transmission scheme. In NRZ, a positive voltage could represent a 1, and a negative voltage could represent a 0.

An NRZ signal is sent down a line, and a receiving station samples the signal at regular intervals to convert the signal back into bits.

The signal will not look exactly like the signal that was originally sent. It will have been distorted and attenuated by the channel. The receiver must map the received signal to the closest symbols.

NRZ is normally not used by itself in practice. More complex schemes can convert bits to signals that better meet engineering needs. These schemes are called **line codes**.

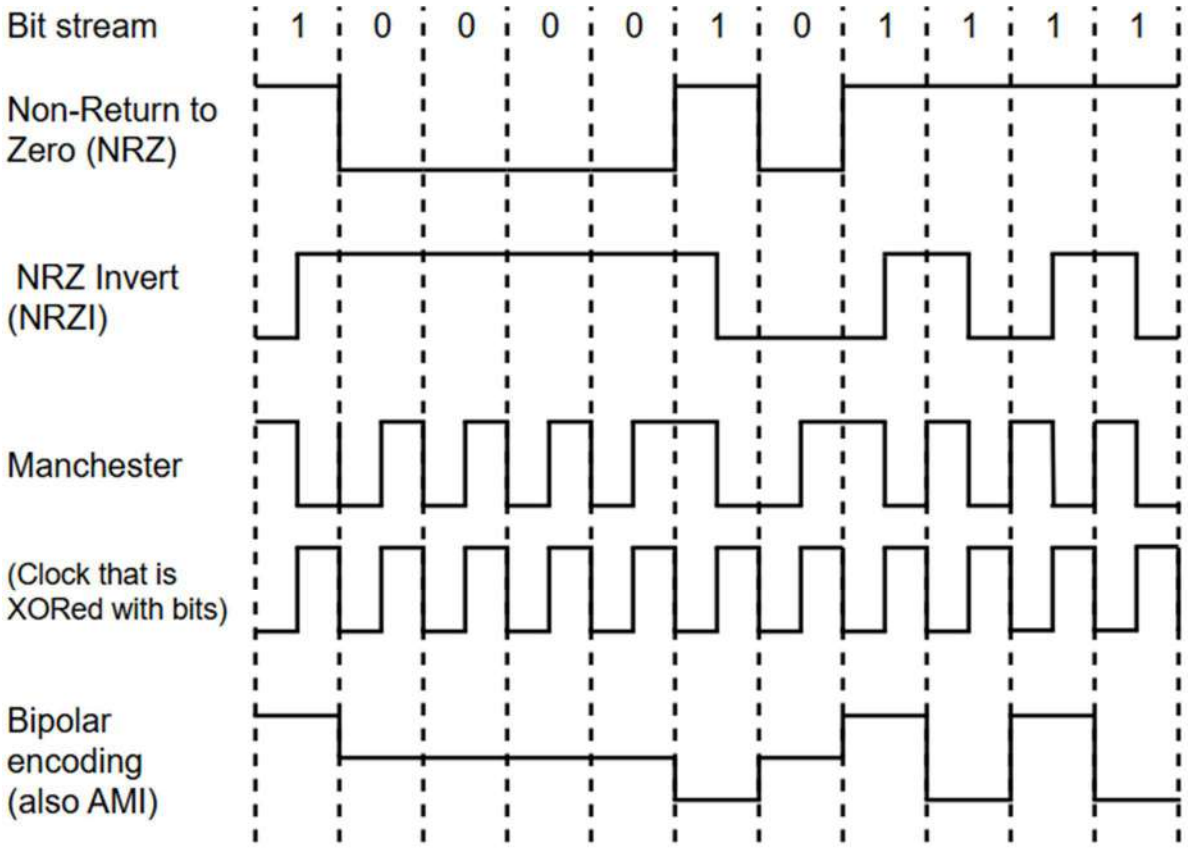


Figure 1: Different line codes

The engineering needs for transmission schemes are:

- Bandwidth Efficiency
- Clock recovery
- Balanced signals

Bandwidth Efficiency

The NRZ signal can cycle between positive and negative levels every 2 bits at a maximum. This requires a bandwidth of at least B/2, where B is bits/second. This is a fundamental limit that NRZ cannot run faster than (without using more bandwidth)

One strategy for using bandwidth more efficiently is to use more than two signalling levels. You could send 2 bits at once as a single **symbol**. This would require four different symbols, and four different levels. This would work as long as the receiver can distinguish the four levels. With this, the rate that the signal changes is half the bit rate, so the required bandwidth is reduced.

The rate at which the signal changes is called the **signal rate**. This is different from bit rate. The bit rate is symbol rate × bits per symbol. Another name for symbol rate is the **baud rate**.

A good scheme maximizes bandwidth efficiency.

Clock recovery

In order for a receiver to read bits from a signal, it must know when one symbol ends and another begins. In NRZ, a long series of 1s or 0s would leave the signal unchanged, so the only way to tell how many bits were sent would be for the receiver to have a very accurate clock.

Accurate clocks are too expensive a solution for commodity devices. The alternative is to send a clock signal to the receiver.

Instead of sending a separate signal for the clock, the clock signal can be mixed with the data signal by XORing the two together. The clock makes a transition in every bit time and it runs at twice the bit rate. When the clock is XORed with 0 it makes a low-to-high transition, which is the clock. When it is XORed with 1 it makes a high to low transition, which is the inverse of the clock. This scheme is called **Manchester encoding** (see Figure 1). Manchester encoding requires twice as much bandwidth as NRZ encoding, because of the clock.

Another approach is to code the data to ensure there are enough transitions in the signal. NRZ will only have a clock problem for long runs of 0s and 1s, so avoiding long runs make it possible for the receiver to stay synchronized.

One approach is to encode 1 as a transition, and a 0 as no transition. This is called **NRZI (Non-Return-to-Zero Inverted)**. The USB standard uses NRZI. NRZI fixes the problem for long runs of 1s, but long runs of 0s will still cause a problem.

Another approach is to map small groups of bits to be transmitted so that groups with successive 0s are mapped to longer patterns that don't have too many consecutive 0s.

4B/5B is a common code for mappings bits. 4 bits are mapped to a 5 bit pattern. This scheme adds 25% overhead. In 4B/5B there are more patterns than there are 4 bit blocks that are mapped in. This means there a free symbols that can be used by protocols as control symbols, for example an idle symbol.

Data (4B)	Codeword (5B)	Data (4B)	Codeword (5B)
0000	11110	1000	11110
0001	11110	1001	11110
0010	11110	1010	11110
0011	11110	1011	11110

Data (4B)	Codeword (5B)	Data (4B)	Codeword (5B)
0100	11110	1100	11110
0101	11110	1101	11110
0110	11110	1110	11110
0111	11110	1111	11110

Another approach to avoiding runs of 0s or 1s is to make the data look random by **scrambling**. A scrambler XORs data with a pseudorandom sequence before the data is transmitted. The receiver then XORs the incoming data with the same pseudo random sequence.

Scrambling doesn't add any bandwidth or time overhead, but scrambling doesn't also doesn't guarantee that there won't be long runs of 0s (although it is unlikely).

Balanced signals

Balanced signals are signals that have as much positive voltage as negative voltage, even over short periods of time.

Balanced electrical signals have no DC component. This is good for media like coaxial cables, which strongly attenuate a DC component. Thus the aim of a good line code should be a balanced signal.

One way to balance a signal is to have a logical 1 represented by alternating between two voltage levels (e.g. +1 and -1), and having 0 be represented as 0. This scheme is called **bipolar encoding**.

Another approach is to use a line code, like 4B/5B. A common code is the 8B/10B line code, which maps bits to balanced symbols. Since there are not enough balanced 10-bit symbols for all 8-bit permutations, not all symbols in in 8B/10B are balanced.

To solve this problem, some input patterns are matched to two symbols: one symbol with an extra 1, and one symbol with an extra 0. In order to keep the signal balanced, the encoder must remember the disparity, next time that it uses an unbalanced symbol, it will choose the symbol that reduces the disparity.

Passband transmission

You can take a baseband signal that occupies 0 to B Hz and shift it up to occupy a **passband** of S to S+B Hz. At the receiver, the signal can be shifted back down to baseband. This type of transmission is called passband transmission "because an arbitrary band of frequencies is used to pass the signal".

With passband transmission, digital modulation is accomplished by modulating a carrier signal that sits in the passband. You can modulate the amplitude, frequency, or phase of the signal.

In **ASK (Amplitude Shift Keying)**, two different amplitudes are used to represent 0 and 1. Multiple amplitudes can be used to represent more symbols.

In **FSK (Frequency Shift Keying)**, two or more different frequencies are used to represent symbols.

In **PSK (Phase Shift Keying)** the carrier wave is shifted 0 or 180 degrees each time time a symbol changes. Because there are two phases, it's sometimes called **BPSK (Binary Phase Shift Keying)**.

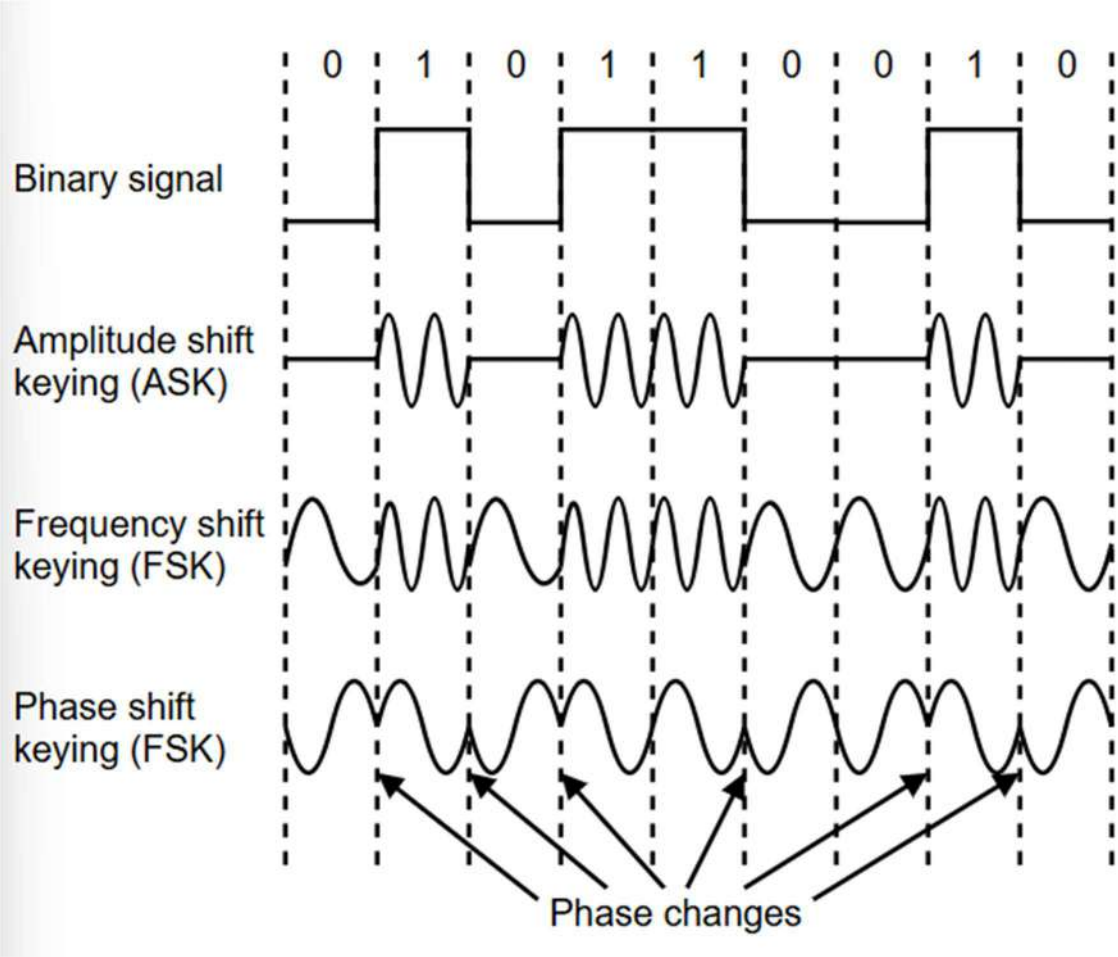


Figure: Modulation schemes

These schemes can be combined together, although frequency and phase cannot. Normally it is amplitude and phase that are used in combination.

Frequency division multiplexing

FDM (Frequency Division Multiplexing) uses passband transmission to share a channel.

FDM divides the spectrum into **frequency bands**. Each user has exclusive possession of a band that they can use to send their signal.

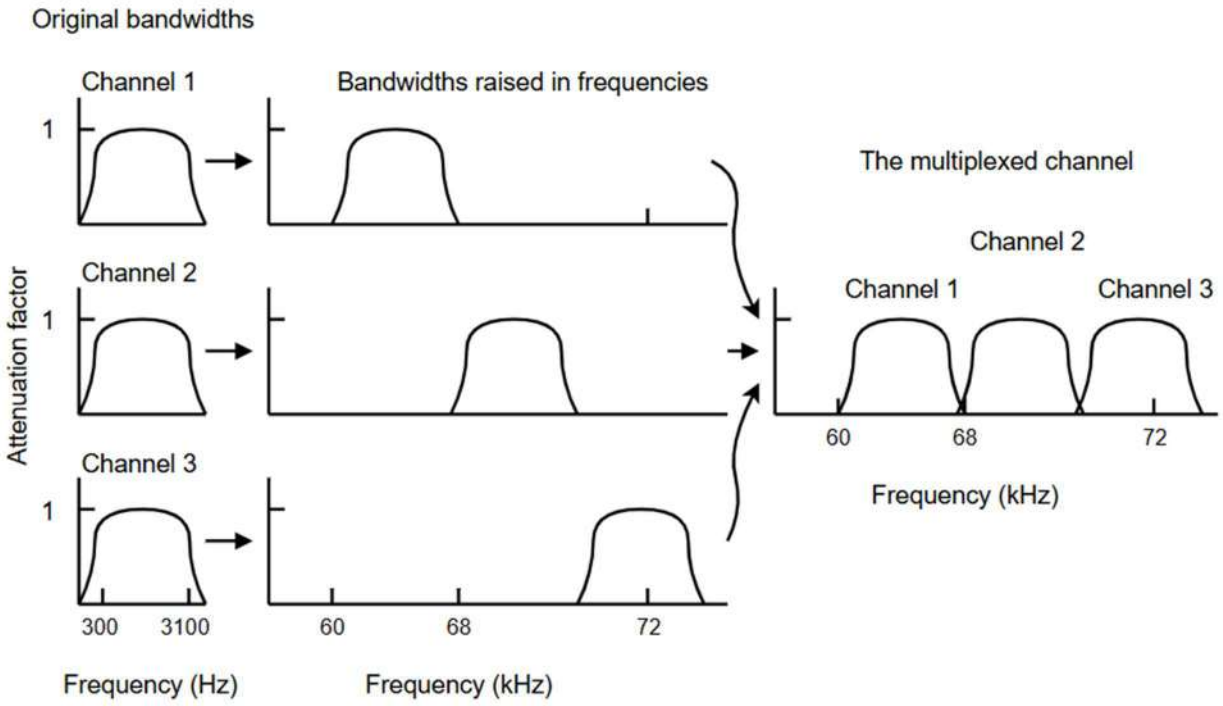


Figure: Frequency division multiplexing

A **guard band** keeps the channels separated.

OFDM

Digital data can be divided without using guard bands. In **OFDM (Orthogonal Frequency Division Multiplexing)** the channel bandwidth is divided into subcarriers, which independently send data.

Signals from each carrier extend into adjacent channels. However, "the frequency response of each subcarrier is designed so that 0 is at the center of the adjacent subcarriers". So subcarriers can be sampled from their center frequencies, without any interference from their neighbors.

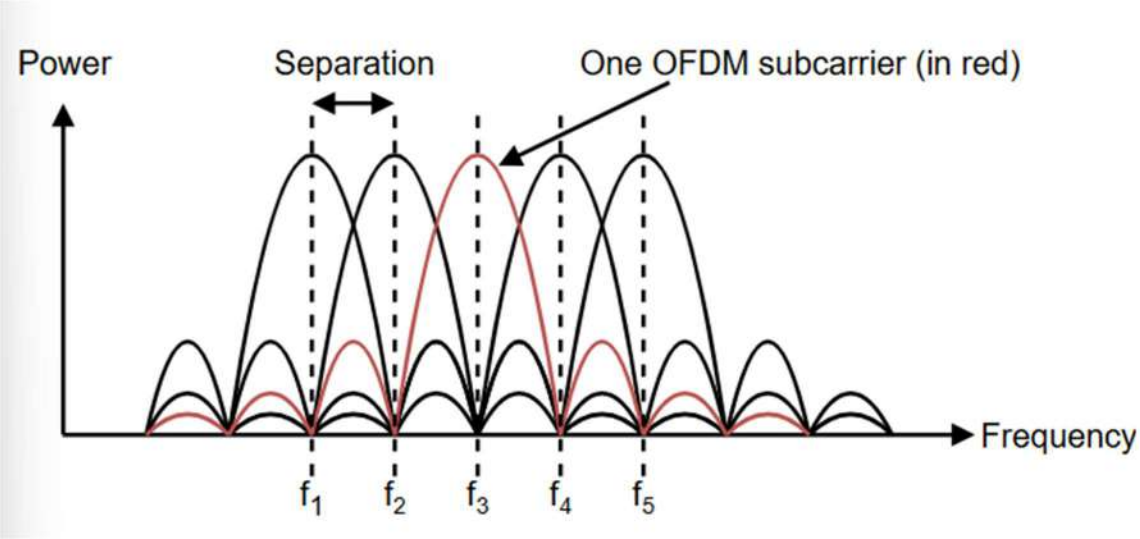


Figure: Orthogonal frequency division multiplexing (OFDM)

Time division multiplexing

TDM (Time Division Multiplexing) works by enabling users full access to the bandwidth for short periods of time.

In TDM, users take turns round-robin style to get the entire bandwidth. Bits from each input are taken on a fixed time slot, and then output to the aggregate stream. The stream runs at the sum of each individual stream.

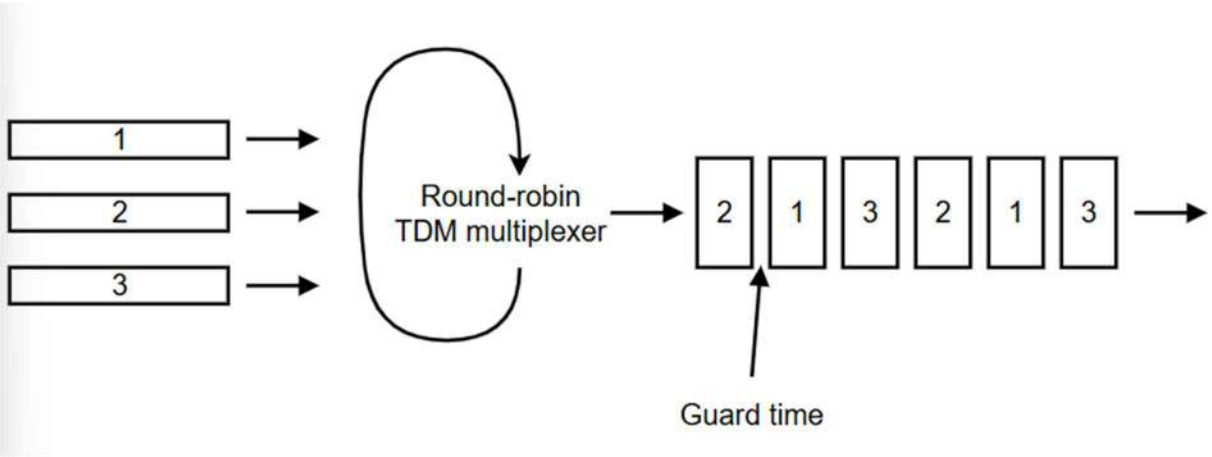


Figure: Time division multiplexing (TDM)

Guard time must be added to accommodate timing variations.

TDM is used in cellular and telephone networks.

Code division multiplexing

CDM (Code Division Multiplexing) is a form of **spread spectrum** communication, where a narrowband signal is spread over a wider frequency band. This allows multiple users to share the same frequency band.

CDM is commonly called **CDMA (Code Division Multiple Access)**. CDMA allows stations to transmit across the entire frequency spectrum all the time. The multiple simultaneous transmissions are separated using coding theory. For CDMA to work, the receiving station must be able to extract desired signals.

In CDMA, each bit time is divided into n short intervals (known as chips). Normally there are 64 or 128 chips per bit.

To transmit a 1, a station sends its chip sequence. To transmit a 0, a station sends the negation of its chip sequence. When multiple stations transmit at the same time, their bipolar sequences add linearly. For example, if during 1 bit period, 3 stations output +1 and 1 station outputs -1, +2 would be received.

The receiving station recovers the original code by computing the normalized inner product of the received chip sequence and the chip sequence of the station whose bit stream it is trying to recover. "If the received chip sequence is S and the receiver is trying to listen to a station whose chip sequence is C, it just computes the normalized inner product, $S \times C$"

This works because pairs of chip sequences are orthogonal (meaning that they cancel each other out).