Networks and Data Communications

The 21st century is typified by devices that are always connected, often wirelessly, to each other and the global Internet. It is common to see the global Internet, and even smaller networks, represented as clouds. A better representation might be ducks. On top of the water, the duck floats along nice and easy, but a lot of paddling happens underneath. Similarly, modern networks have a lot of interconnections and moving parts. Information technology professionals are expected to keep that duck floating serenely along. Effective network problem solving requires a clear mental model of the interconnections and moving parts.

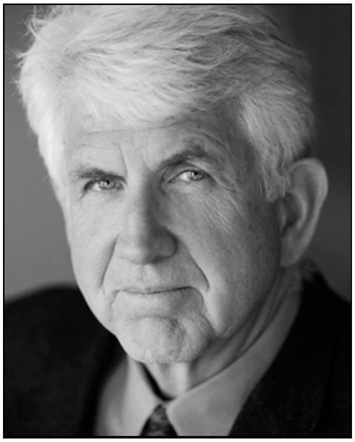


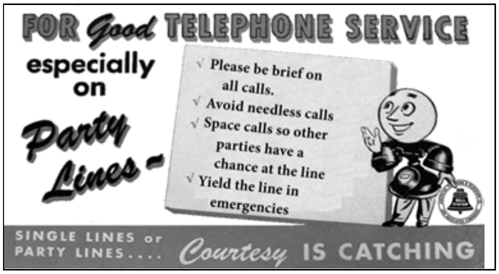
Figure 0-1
Bob Metcalfe, inventor of Ethernet and formulator of Metcalfe's Law

The Internet is so named because it is an **internetwork**, an interconnection of many local area networks. This chapter explores network technology, the mechanism by which networks operate, and the techniques for interconnecting disparate networks.

We expect our gear to communicate seamlessly and, for the most part, it does. This is because of the development of standards, the importance of which was covered in Chapter 0. A study of data communications and networks is largely a study of the relevant standards.

0.1 Messages, Protocols, and Channels

At the simplest level, data communications consists of messages, protocols, and channels. Messages are exchanged between or among hosts, also called nodes, which are computers equipped with the necessary hardware and software for communication. More specifically, a **node** is any device attached to a network while a **host** is a general-purpose computer. A **message** might be an operating system update, a streaming video, a music download, or digital voice on its way to a grandmother's landline.



FOR *Good* TELEPHONE SERVICE

especially on *Party Lines* –

√ Please be brief on all calls.

√ Avoid needless calls

√ Space calls so other parties have a chance at the line

√ Yield the line in emergencies

SINGLE LINES or PARTY LINES.... *Courtesy* IS CATCHING

Figure 0-2
Protocol for party line telephones

Protocols are rules. In data communications, they are rules for communication. The same word is used in diplomacy. Two hosts that share the same protocols and have suitable network interface control (NIC) hardware can communicate with each other. Figure 0-2 shows the protocol for party line telephone service.[1] Protocols are established by standard-setting organizations, among them the Internet Engineering Task Force (IETF), the International Organization for Standardization (ISO), and the International Electrotechnical Commission (IEC.) Protocols apply to the hosts at the ends of a channel and to the communication channel itself. They also govern **intermediate nodes**, which are specialized computing devices for routing or switching signals.

The communication **channel** is the mechanism by which signals are transmitted from one network node to another. The connection between two nodes is called a **link**. In network diagrams, a link is often represented as a lightning bolt or a pipe. The mechanism might be radio, fiber optic cable, or copper cable. It is likely that more than one communication channel is involved with a connection to a remote resource. A mobile device's connection to an Internet web site will involve cellular radio and probably both copper and fiber optic connections. The **bandwidth** of a channel is the theoretical amount of information the channel can carry per second, measured in Hertz. A channel may carry more than one link. The **speed** of a link is the transmission rate, measured in bits per second. Modern networks operate at speeds measured in megabits, millions of

[1] Up to the middle of the 20th century, it was not uncommon for two or more families to share a single telephone line, especially in rural areas. One had to listen to be sure the line was not in use before attempting to place a call. That is remarkably similar to how Ethernet works in a non-switched environment.

bits per second, or gigabits, billions of bits per second. **Latency** is the delay in transmission as signals pass through the network from source to destination.

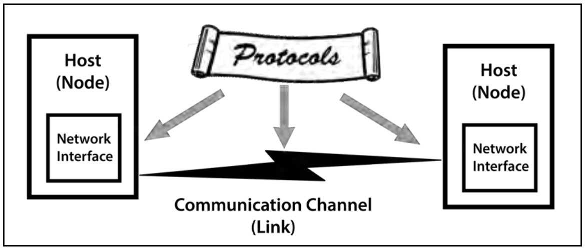


Figure 0-3
Abstract diagram of the communication process

Figure 0-3 is a high-level diagram of a data connection, showing two nodes, their network interface hardware, and the connection between them. Protocols and standards govern all parts of the process, from the signaling mechanism and shapes of the connectors to the format of the messages exchanged.

Circuit switching and packet switching

To place a telephone call in 1925, one lifted the receiver of a landline telephone instrument. That caused an electric current to flow and alerted a telephone operator, who said “Number, please?” You spoke the number you wanted. If the destination telephone wasn’t already in use, the operator physically connected the wires from your phone to those of the person you called using plugs and sockets. The connection remained in place until you hung up the telephone instrument. Because of the need for an end-to-end connection, long distance calls were complex and expensive. A ten-minute call from New York to San Francisco in 1925 cost $27.75. (Waldon & Lande, 1997) That’s more than $400 in 2020 dollars.

Figure 0-4
A telephone switchboard
Seattle Municipal Archives

Now consider mailing a letter from New York to San Francisco. In 1925, that cost two cents for a letter of up to one ounce, the equivalent of about thirty cents today. You had an expectation that the letter would be delivered but little or no idea of the path it would take. A little thought would convince you that many other letters traveled along with yours on various parts of that journey, and that a second letter mailed to the same person might take a somewhat different path.

The telephone call example illustrates a **circuit-switched** network.[2] When a call is placed, a connection is established between caller and called party. It remains active for the duration of the call. Circuit switching is **connection-oriented** and reserves the full bandwidth of the channel even when it is not carrying information, and so introduces inefficiency. While the connection is established, no other connection can be made. If computer connections were circuit-switched, one person streaming a video would lock out everyone else in the household until the video ended.

The letter example illustrates **packet switching**. Packet switching divides messages into packets of perhaps a few hundred bytes. Each packet is transmitted independently. This allows use of the available bandwidth to be maximized. It also allows packets from different messages to be interleaved on the same channel. One person might be streaming a video while another is browsing a news site with neither aware of the other's activity. Packet switching is **connectionless.** Modern data communications systems employ packet switching. Although the communication hardware is no longer circuit-switched, we will see the concept of persistent connections at a higher level of abstraction as we explore communication protocols, particularly the transmission control protocol, TCP.

"Reliable" and "unreliable" protocols

The words "reliable" and "unreliable" are used in a way that's different from their common meanings when describing data communications protocols. An analogy with the postal service may help. If you place a properly addressed letter in the mail, it is nearly certain to be delivered, but the postal service doesn't confirm the delivery. That is an unreliable mechanism. If you pay extra for certified mail with a return receipt, you are nearly certain to get a green post card

[2] In 21st century landline calls, only the connection between the telephone instrument and the telephone central office is circuit switched. Until the middle of the 20th century, landline calls were circuit-switched end-to-end.

in the mail confirming that your mail piece was delivered. Note that your original mail piece can still be lost, or the return receipt card can be lost. If, after a few days, you don't get that green post card, you know that something is wrong. You don't know what, but you know there's a problem. That is a reliable mechanism.

We can now provide a definition. A **reliable protocol** acknowledges receipt of messages and an **unreliable protocol** does not. In a reliable protocol, absence of an acknowledgement is an error which is handled by having the sending node retransmit the unacknowledged message(s). The receiving node must identify and discard duplicate messages. A duplicate message results when it is the acknowledgement and not the original message that is lost.

Local and wide area networks

At the beginning of the 1970s, when researchers began to get serious about data networks, where was a clear distinction between local area networks, called LANs, and wide area networks, called WANs. In **local area networks**, the operators of the networks owned the wires that connected the hosts on the network together. So, LANs connected devices in one department, or a single building, or maybe a college or corporate campus. To connect to a location across town, one need the help of a telephone company. Telephone service was mostly a monopoly then, dominated by AT&T. Connections were expensive and speeds were very limited. Those who operated networks soon realized that it was far more effective to connect two networks than two devices when telephone company lines were involved. So, "**wide area network**" even then was a misnomer; people were building internetworks, in the same fashion as the global Internet but on smaller scales.

Things are a little less clear today. There are still local area networks where the network owners also own the wires and wireless access points. There are still private internetworks operated by entities like banks and governments, but many organizations use the Internet as a mechanism to connect local area networks.

Your cell phone isn't a local area network, but it isn't an internetwork, either. Today we talk about local area networks, Internet service providers, and "**endpoints**," which can include a cell phone, a wired desktop computer, and a laptop with wireless access to a local wireless access point.

0.2 Protocol Stacks and the OSI Model

The concept of a **protocol stack** represents modularization of the hardware and software functions necessary for data communication. Not only is modularization good programming practice, in the case of data communications systems, it means that one part of the process can be replaced without disrupting everything else. For example, the transmission medium could be changed from copper wire to cellular radio without changing the way application programs deal with the network. There are two protocol stacks of interest to the student of information technology, the Open Systems Interconnection (OSI) model and the Internet Protocol Suite model. The Internet Protocol Suite model is often called the TCP/IP model because the transmission control protocol (TCP) and the Internet Protocol (IP) are the fundamental protocols of the model.

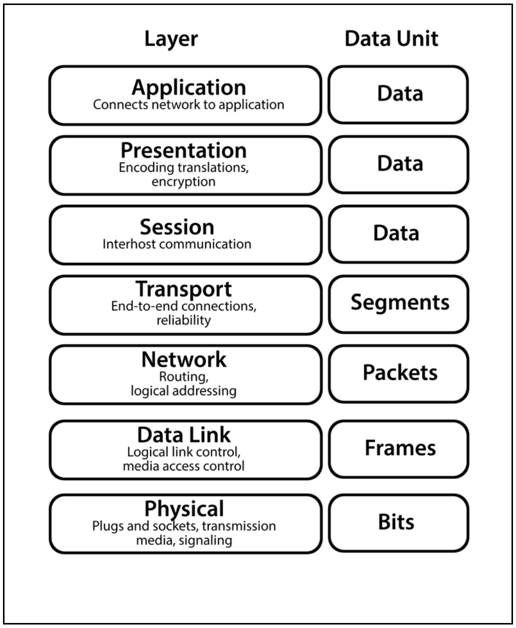
0.2.1 The OSI Model

The **OSI model** is a reference model, that is, an abstract model describing interaction with a network. It does not describe any physical or programming interfaces. For that reason, the OSI model is of primarily of theoretical interest. It provides arguably the cleanest separation of the various required functions of a protocol suite. More importantly, it includes the physical layer, which is subsumed in the link layer of the TCP/IP protocol suite. The seven-layer OSI model is shown in Figure 0-5.[3]

The bottom-most layer, the **physical layer**, is concerned with the physical connectors, the transmission medium, and the signaling protocol. Its job is to move bits from one network node to another.

[3] The names of the layers are something to look up, not to memorize. If you should ever have to memorize the layer names, the mnemonic, from bottom to top, is, "People Do Not Throw Sausage Pizza Away."

The **data link layer** organizes bits into frames, the logical unit of transmission. The data link layer is responsible for logical link control and media access control. The media access control sublayer performs media arbitration, that is, deciding which nodes may transmit, and frame synchronization, that is, dividing the bit stream from the physical layer into frames. The logical link control sublayer is responsible for addressing and may provide flow control and error control.



Layer	Data Unit
Application Connects network to application	Data
Presentation Encoding translations, encryption	Data
Session Interhost communication	Data
Transport End-to-end connections, reliability	Segments
Network Routing, logical addressing	Packets
Data Link Logical link control, media access control	Frames
Physical Plugs and sockets, transmission media, signaling	Bits

Figure 0-5
The OSI protocol stack

The **network layer** provides for logical addressing, for example, using IP addresses. It also provides routing between networks. The network layer is connectionless in the sense that its job is transporting individual packets.

The **transport layer** is responsible for segmenting messages, for example, cutting a large video download into smaller segments for transmission, then reassembling them at the other end. It is responsible for error control and flow control. **Flow control** means signaling a sending node to pause if data are being sent faster than they can be accepted.

The **session layer** controls exchanges of related messages between two network endpoints. The session layer establishes a connection, facilitates exchange of messages, and closes the connection at the end of the session.

The job of the **presentation layer** is data independence. For example, one end of a network connection may use double-byte UCS-2 characters while the other uses UTF-8. The presentation layer performs the necessary translation. Compressing and decompressing data and encrypting and decrypting also happen in the presentation layer.

The **application layer** provides the interface between the network and the application. The application could be a web browser or chat program, or it could be an internal application at a bank to move data between branches. The application program communicates with the application layer of the network protocol stack using an application program interface (API) to send and receive data. The application program does not need to be aware of any of the other parts of the network protocol stack. In fact, insulating applications from the details of networking is the reason for the concept of a protocol stack. From the viewpoint of the application developer, the application programs at the two ends of the network connection are communicating directly with one another using those API calls.

0.3 Ethernet, Local Networks, and the Data Link Layer

A general-purpose local area network of the 21st century is nearly always an Ethernet network, often running at gigabit speeds.[4] A local area network serves a home, a department, a building, or perhaps a campus, and connects to other networks such as the global Internet through the use of routers.

Bob Metcalfe invented Ethernet in 1973 at Xerox Corporation's Palo Alto Research Center (PARC). Xerox filed a patent application for Ethernet in 1975. In 1976, Metcalfe and David Boggs published a paper describing Ethernet in *Communications of the ACM*. (Metcalfe & Boggs, 1976) Xerox Corporation, together with Digital Equipment Corporation and Intel developed Ethernet into an open standard in 1980, and it was adopted as an Institute of Electrical and Electronics Engineers (IEEE) standard in 1985. Ethernet is named for *luminiferous ether*.[5]

The original design for Ethernet was based on a shared transmission medium, with all nodes connected to a single, long and thick coaxial data cable. A transmission by any node on the network was received by all other nodes, but was ignored by all but the node to which it was addressed. The icon still used to

[4] Specialized networks using technologies other than Ethernet are found in automotive, aviation, and manufacturing settings, among others.

[5] Luminiferous ether was the hypothetical substance supposed to transmit light in the way that air transmits sound. In 1905, Einstein's special theory of relativity showed that it was not necessary to assume the existence of luminiferous ether to explain the propagation of electromagnetic waves.

mark wired Ethernet connections depicts three computers connected to a common cable.

The thick coaxial cable was cumbersome, difficult to install, and difficult to maintain. A fault anywhere on the cable could bring down the whole network. Development quickly progressed to the use of twisted-pair cable connected with devices called hubs, and by 1994, to the use of fiber optic cable.

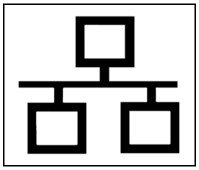


Figure 0-6
The Ethernet icon

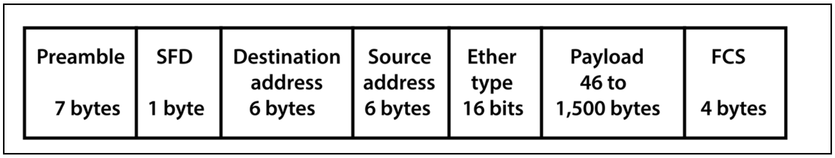
Today, Ethernet can run at speeds of up to 400 gigabits per second. Gigabit Ethernet over twisted pair cable is common and inexpensive. Wireless Ethernet, called Wi-Fi, is also common.

0.3.1 The Ethernet Data Link Layer

Every Ethernet **network interface controller** (NIC) has a 48-bit address assigned when the device is manufactured and called the **media access control (MAC) address**. The first three bytes are assigned to the device manufacturer by the IEEE and are called organizationally unique addresses. Two bits of the first byte are reserved and are always zero in the prefixes assigned to manufacturers, leaving 22 bits, enough to assign four million organizationally unique addresses. Large manufacturers have more than one organizationally unique address. The remaining three bytes are assigned during manufacture such that the resulting address is globally unique.

MAC addresses are written as six hexadecimal pairs, separated by colons. An example is 00:11:25:d3:12:77.

An Ethernet NIC will receive all traffic on its part of the network, but ignore data not addressed to it. Exceptions are the broadcast address, 48 bits of all ones, which is acted upon by every NIC, and a possibly configurable list of **multicast addresses** that allow a single message to be processed by more than one NIC. A NIC may also be placed in **promiscuous mode**, which allows it to accept and process all traffic it receives. This is useful for network monitoring applications.



Preamble	SFD	Destination address	Source address	Ether type	Payload	FCS
7 bytes	1 byte	6 bytes	6 bytes	16 bits	46 to 1,500 bytes	4 bytes

Figure 0-7
An Ethernet frame

The unit of transmission in Ethernet is the frame. Figure 0-7 shows an Ethernet **frame**. The preamble is 56 bits of alternating one and zero. This allows the sending and receiving NICs to synchronize bit timing. The one-byte start of frame delimiter (SFD) provides for byte synchronization and marks the start of the meaningful data in the frame. It is followed by the destination and source addresses, the Ether type field, the payload data being transmitted, and a frame check sequence (FCS) value. The FCS is a cyclic redundancy check. It is computed by the sending NIC and verified by the receiver. A mismatch indicates that a transmission error has occurred. In that case, the receiving NIC discards the frame. It is up to a higher-layer protocol to recognize that a frame is missing and request retransmission.

An alternate IEEE standard allows the field labeled Ether type in Figure 0-7 to contain the payload length. To allow the two uses of the field to be distinguished, the Ether type field must be greater than 1,535.

Non-standard jumbo frames can carry payloads larger than 1,500 bytes. This has the advantage of reducing the frame overhead caused by the preamble, SFD, and header information. The trade-off is that larger frames may have an adverse impact on latency.

0.3.2 Collisions and Switching

In the original design of Ethernet, all nodes were connected either to a single data cable or to each other through a hub.. That arrangement is called **multiple access**. The protocol requires that each node check that the cable was idle, *i.e.* that no other node was transmitting, before beginning a transmission. Checking whether another node was transmitting is called **carrier sense**. If another node was transmitting, a node with a frame to send had to wait until the cable was idle. With this arrangement, it was still possible for two nodes to transmit at once if both detected that the cable was idle simultaneously. That circumstance was called a **collision**, and Ethernet NICs were designed to be able to detect collisions. When two transmitting nodes detected a collision, both stopped, waited a random time, then listened for the cable to be idle. Since it was very unlikely that the two random times would be equal, one node would "win" and begin to transmit. The other would have to wait until the cable again became idle. The collection of nodes for which mutual collisions is possible is called a **collision domain**.

The combination of carrier sense, multiple access, and collision detection is abbreviated **CSMA/CD**.

The CSMA/CD design worked well for small networks and low traffic volumes. As network nodes and traffic increase, the number of collisions increases. At some point more time is spent recovering from collisions than transmitting data and network throughput degrades badly.

Switched Ethernet

The solution to the problem of collisions degrading network throughput is to remove the concept of collision domains using **Ethernet switches**. In a wired or optical Ethernet, each node is connected to a port on the switch. Both the node and the switch can transmit and receive simultaneously, an arrangement called **full duplex**. In such an arrangement, no collisions are possible because there is no longer a shared medium; communication is between a single host or node and a port on the switch.

Switching applies only to wired and optical Ethernet. Collisions remain possible in wireless Ethernet when more than one wireless device is using one wireless access point. Wi-Fi uses a technique called **collision avoidance** to minimize the number of collisions.

In normal operation, an Ethernet switch receives a frame on one port and transmits it again on a different port based on the destination address in the frame. When an Ethernet switch is first started, it has no information about which ports are associated with what MAC addresses. In that case, it sends the frame on all ports except the one on which it came in. As frames are received, firmware or software in the switch records which MAC addresses are on what ports, something like “00:11:25:d3:12:77 is on port 16.” Within a very short time, the switch is able to direct most frames to a single destination port rather than all ports. Broadcasting is still necessary from time to time as nodes join or leave the network.

A major drawback of a switched Ethernet network is the potential for a switching loop. If two ports on a switch are connected to each other, or there are two connections between switches, such a loop exists. If a frame for which there is no entry is received, it will be broadcast to every port except the one on which it arrived. If two ports are connected together, each will receive and rebroadcast the frame continuously. That is called a **broadcast storm**, and it will render the network useless until it is stopped, possibly by resetting the switch. However, another broadcast packet will cause another broadcast storm.

The spanning tree protocol and shortest path bridging protocol

Loops that can cause broadcast storms can be introduced accidentally or even deliberately as a way to provide redundancy. The invention of the **spanning tree algorithm** by Radia Perlman in 1985 allowed software or firmware in switches to find those paths through the network sufficient to reach every node, that is, to span the network. Switch ports unnecessary to the spanning tree are disabled and loops cannot exist. A change in the topology of the network, such as by a connection failing, causes the spanning tree to be recomputed. That makes redundant connections possible, but only one connection of a redundant set can be

Algorhyme

I think that I shall never see
A graph more lovely than a tree.

A tree whose crucial property
Is loop-free connectivity.

A tree which must be sure to span
So packets can reach every LAN.

First the root must be selected.
By ID it is elected.

Least cost paths from root are traced.
In the tree these paths are placed.

A mesh is made by folks like me
Then bridges find a spanning tree.

Figure 0-8
Algorhyme *by Radia Perlman*
Copyright © 1985 by ACM

used at a time. Perlman's description of the spanning tree algorithm is in Figure 0-8. (Perlman, 1985)

The **shortest path bridging protocol** was developed by the IEEE 802 working group in 2006 and finalized in 2012. SPB allows traffic on redundant links thereby maximizing network throughput while preventing or mitigating loops. As of this writing, SPB is found only in enterprise and service provider grade equipment.

0.3.3 The Ethernet Physical Layer

Standardization at the physical layer as well as the data link layer means you can buy a computer from one manufacturer, an Ethernet switch from another, and cable from a third and expect them to work together seamlessly provided they are for the same physical specification.

The physical specification for Ethernet is a three- to five-part description consisting of speed, modulation technique, physical medium, encoding method, and either number of lanes for local connections or maximum distance for wide-area connections. For example, gigabit Ethernet has a physical specification of 1000BASE-T. The 1000 is the speed in millions of bits per second, BASE indicates baseband, *i.e.* no other carrier signals on the cable, and -T says the transmission medium is category five twisted pair cable.

The Ethernet physical layer for ten and 100 megabit speeds describes a medium-dependent interface (MDI) which connects the transmitter of one device to the receiver of the other, and vice-versa. End devices like computers are generally MDI devices and central devices like switch are generally MDI/X devices with transmission and reception lines "crossed over." Modern 10/100 megabit devices and all gigabit and higher devices negotiate the "crossover" automatically when the cable is connected.

Modern consumer-grade and small-office Ethernet switches automatically negotiate to the highest speed common to the two devices for 100 megabit and gigabit speeds, and sometimes for 10/100/1000 speeds.

Wireless Ethernet

Devices participating in a wireless Ethernet are called **stations**. Wireless Ethernet uses radio signals to transfer information. For that reason, it must still operate in the carrier sense multiple access (CSMA) mode because there is no concept of a private connection to a switching device. Wireless Ethernet, or Wi-Fi, attempts to avoid collisions using the same listen-before-sending mechanism as the original wired Ethernet. In addition, a station may optionally send a request to send (RTS) control frame. The access point or controlling station will reply with a clear to send (CTS) control frame, reserving the radio channel for the original station to transmit a data frame.

Wireless Ethernet is standardized by the IEEE as a set of numbered 802.11 and suffixed by letters. The Wi-Fi Alliance has published a set of version numbers, Wi-Fi 1 to Wi-Fi 6 which they believe to be more consumer friendly. In the United States, most wireless Ethernet equipment operates in the 2.4 gigahertz (GHz) or 5 GHz frequency bands. In 2020, the U.S. Federal Communications Commission approved a frequency allocation in the 6 GHz band for wireless Ethernet. There are also frequency allocations in the 60 GHz band, intended for wireless audio-visual applications. Although higher frequencies allow for higher data rates, they are more easily blocked by walls and other obstructions.

Particularly for wireless networking, security from eavesdropping is a concern. When a signal is transmitted by radio, not only Wi-Fi, but also Bluetooth or cellular, any suitable receiver within range can pick up the signal. To prevent eavesdropping, such signals are nearly always encrypted. The wireless Ethernet standards include a provision for encryption and for authentication with an access point. Encryption is discussed further in Chapter 7. ****

0.4 The Internet Protocol Suite – TCP/IP

The TCP/IP protocol suite is the protocol suite of the global Internet. It was specifically designed to be independent of the physical connection. A joke at the time was that it could run over two tin cans and a string. The TCP/IP model is a four-layer model, with the underlying physical and data link layers assumed, but not a part of the model. The Internet Protocol Suite model is shown in Figure 0-9.

The lowest level is the **link layer**. The link layer includes the address resolution protocol (ARP), which translates Internet Protocol addresses to the addresses used by the data link layer of the OSI model. Everything else is assumed to exist and is not explicitly specified in the protocol. The scope of the link layer is all hosts directly accessible on the local network. The data units are those of the local network, often Ethernet frames.

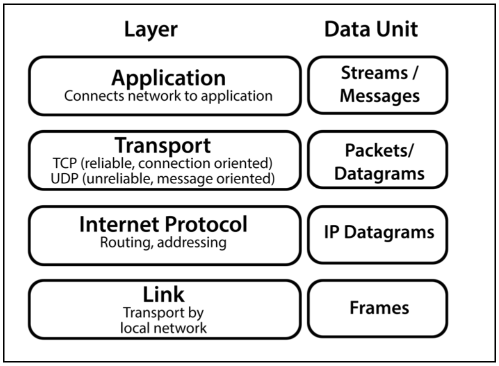


Figure 0-9
The Internet suite protocol stack

The **address resolution protocol** (ARP) solves the problem of communication on a local area network when a logical address such as an Internet Protocol address is known but the link layer address, *e.g.* a MAC address is not. A node or host wanting to communicate with a node or host with unknown link layer address sends a broadcast message containing the Internet Protocol address. The node with that address replies, and the link layer frame contains that node's link layer address. The correspondence of logical address and link layer address is stored in an **ARP table** for future communications with that host.

The **Internet Protocol layer** is responsible for addressing and routing IP datagrams. Each datagram is an independent message to be transmitted from source to destination. So, a large video download will be divided into many datagrams, each handled separately by the Internet Protocol layer. The Internet Protocol layer does not distinguish among the several protocols of the transport layer.

There are two incompatible addressing schemes at the Internet Protocol layer. IPv4 was the original addressing scheme of the Internet, numbers one to three having been used for experimental protocols. IPv4 is based on 32-bit addresses, which are discussed in Section 0. IPv6 was accepted as a standard in 1996; significant use of IPv6 addresses began in the early 21st century. The 64-bit IPv6 addressing mechanism is discussed in Section 0. IPv5 was a proposed streaming standard that was never officially adopted.

The Internet Control Message Protocol (ICMP) has as its primary purpose exchanging messages among network devices. It is visible to users in the PING and TRACEROUTE applications.

There are three major transport layer protocols, the transmission control protocol (TCP), the user datagram protocol (UDP), and stream control transmission protocol (SCTP). In addition, a protocol developed by Google, QUIC (a name, not an acronym), is on track to be adopted as a standard by the IETF. Strictly speaking, transport layer security (TLS) is implemented on top of the transport layer protocol, and is not itself a transport layer protocol. For the purposes of this book, it will be covered here, along with the transport layer protocols. There are several other Internet suite protocols not discussed here.

Transmission Control Protocol

The transmission control protocol (TCP) is a connection-oriented and reliable protocol. At the application layer, each application appears to be talking directly to the corresponding application at the other host. Communication is through an application program interface, or API. Often the API is a **socket**, an interface specification developed by Berkeley Systems Division and standardized as a part of POSIX, the IEEE portable operating system interface.

A connection is established through a TCP handshake process in which the two communicating hosts exchange sequence numbers. The sequence numbers are assigned randomly to help thwart session hijacking. The sequence numbers are used to re-order packets that may arrive out of order. Each station also acknowledges receipt of the other's packets. The TCP header includes an acknowledgement field so that a separate message is not needed for acknowledgements. Unacknowledged packets are retransmitted. If the problem that caused a retransmission was with the acknowledgement and not the original packet, the sequence number allows the duplicate packet to be identified and discarded. The TCP header contains a checksum that is verified on reception. Packets for which checksum verification fails are discarded and will be retransmitted because they aren't acknowledged.

Flag bits and other fields in the TCP header provide for flow control and optionally for explicit congestion control. One of the flag bits is used for gracefully closing the connection.

In addition to IP addresses for the endpoints, a TCP connection requires port numbers for each endpoint. Port numbers do not refer to any type of physical connection. Instead, they are 16-bit numbers that are used by the operating system to send data to the correct application. A TCP connection is completely characterized by the protocol number, the two IP addresses, and the two port numbers. A program such as a web browser wanting to make a connection to a server asks the operating system to assign a random port number for the outgoing part of the connection. The 65,536 possible port numbers are divided into well-known port numbers, registered port numbers, and dynamic port numbers. The application will usually use a **well-known port** for connection to the destination host. For example, an HTTP connection will be make on port 80, an HTTPS connection on port 443. The Internet Assigned Numbers Authority (IANA) is responsible for assigning well-known ports, and most of the port numbers below 1024 have been assigned. Many higher-numbered ports are also used, registered to particular applications by the IANA. The destination server must be configured to listen on a particular port.

Port numbers 49,152 to 65,535 are available for operating systems to assign dynamically.

User Datagram Protocol

The user datagram protocol (UDP) is connectionless and unreliable. It provides for individual messages, called datagrams, which are not acknowledged at the protocol level. Since each datagram is handled separately, it is possible for datagrams to arrive out of order.

Because there is no handshake at the beginning of communication and no need for acknowledgements, UDP is considerably faster for short messages than

TCP. It is particularly well-suited for query / response applications such as domain name system queries, discussed beginning on page 21. UDP is also used for streaming applications such as voice over IP, where the occasional loss of a packet is not critically important.

UDP uses port numbers to direct messages to particular applications. Since UDP is a different protocol from TCP, the port numbers are disjoint. That is, a host may use a port number for TCP and the same number for UDP without conflict.

The theoretical maximum size of a UDP datagram is 65,536 bytes minus the length of the header, so usually 65,507 bytes for IPv4. However, large datagrams are likely to be fragmented at the IP layer. Since there is no error handling, the loss of one fragment is effectively the loss of the entire datagram. For that reason, many applications using UDP limit datagram size to around 512 bytes. That is small enough to avoid fragmentation at the IP level except in very unusual circumstances because the minimum allowed transmission unit for the Internet is 576 bytes.

Stream Control Transmission Protocol

The stream control transmission protocol, SCTP, is a connectionless protocol. Like UDP it transmits messages (datagrams) rather than providing a connection between the endpoints. Unlike UDP, SCTP provides a reliable connection. It includes other features for reliability, such as multi-homing. Multi-homing means one or both endpoints may have multiple IP addresses allowing transparent fail-over in case of a link failure.

Transport Layer Security

Transport layer security (TLS) is not an Internet Protocol transport layer; it operates on top of the transport layer and below the application layer. However, it is important in the context of securing communication between networks and so should be included in this chapter. TLS provides privacy of communications by encrypting the data traveling between endpoints. TLS uses message authentication codes to protect the integrity of communications. Digital certificates provide a measure of assurance of the identity of the server endpoint. Client certificates are optional. Modern implementations provide forward secrecy, making it impossible to decrypt future communications or stored past communications even if the server's private key is compromised. Forward secrecy is discussed further in Chapter 7.

Many web site operators call their web sites "secure" when transmission between browser and web server is encrypted with TLS. However, TLS only secures *transmission* between browser and web server. It is still possible for the server itself to be carelessly managed or inadequately secured.

The QUIC Protocol

At the time this is written, the QUIC[6] protocol is an Internet draft, not yet a standard. Like TLS, QUIC is intermediate between the transport layer and the application layer. It uses UDP for transport. Error recovery, lost datagram recovery, and datagram reordering are performed at the QUIC level. QUIC was devised by Jim Roskind and others at Google to speed up web services by taking advantage of the characteristics of interactions between web servers and web browsers, such as establishing a TLS connection during the initial setup instead of requiring separate TCP and TLS handshake operations. Although not yet a standard, it is supported the servers operated by Google, on the LiteSpeed and Ngnix server packages and by at least on commercial content distribution network. QUIC is supported on the major web browsers.

As of fall, 2020, the Internet draft specification for HTTP/3 includes QUIC as a transport protocol in preference to TCP. Although not yet a standard, HTTP/3 is implemented in the major browsers.

0.4.2 TCP/IP Names and Addresses

IP-based networks use one of two formats of logical addresses. Users of such networks nearly always refer to hosts and network resources by name. they nearly always expect the machines they use to "just work," that is to be configured automatically or to be self-configuring.

[6] It's a name, not an acronym, pronounced *quick*.

Those who design, install, maintain, and troubleshoot IP-based networks must have a thorough understanding of the relationships between names and addresses and of the mechanisms used to configure hosts on the network.

Routers and Firewalls

Routers are switching devices that operate at layer 3, the network layer of the OSI model, which is equivalent to the Internet Protocol layer of the TCP/IP model. The simplest routers for home or small office networks have two interfaces, a LAN interface that operates at the Ethernet data link and physical level, and a WAN interface. The WAN interface might be a cable or DSL modem or an optical interface for connection to an Internet service provider. Its principal purpose isn't really routing, it is to translate from the Ethernet physical and data link layer to the physical and data link protocol needed by the connection to the service provider. Such a device may include an Ethernet in the same physical package and have more than a single Ethernet port, and even a wireless access point.

Enterprise and carrier-grade routers have multiple ports and, like switches, choose a port on which to forward each datagram. The major difference is that routers use logical addresses and make routing decisions based on the network portion of the address as discussed below. The routing information protocol (RIP) and open shortest path first protocol (OSPF) are "interior" routing protocols, suitable for use within a single enterprise. The border gateway protocol (BGP) is an "exterior" suitable for routing among peer systems. A thorough discussion of routing protocols is suitable for a networks course.

A **firewall** serves as a choke point on a network and determines which packets are allowed to pass and which will be dropped. The decision of whether to allow a packed is made according to configurable firewall rules. Such rules may consider the direction, incoming our outgoing, of the packet, source and destination port numbers, and source and destination addresses. Some firewalls engage in packet inspection. Shallow packet inspection involves checking the headers of higher level protocols, which are considered payload at the Internet Protocol level. Some firewalls look at the data portion of the payload. This called deep packet inspection. Deep packet inspection is used to attempt to detect unauthorized data transfer or attempts at unauthorized entry (intrusion) into systems attached to the network.

The Domain Name System (DNS)

The job of the domain name system is to translate human-friendly domain names into logical IP addresses. The domain name system is a hierarchical, distributed database that maps domain names to IP addresses, provides inverse mapping, and provides resource records several other functions. Its principal use is mapping domain names to IP addresses. Name servers receive requests specifying a name and return an IPv4 or IPv6 Internet Protocol address. DNS operates at the application layer of the Internet Protocol suite.

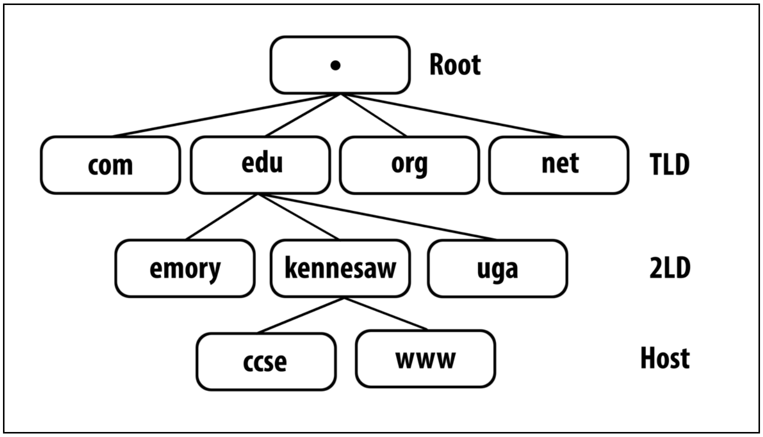


Figure 0-10
The domain name system (DNS) database is organized as a tree structure

Figure 0-10 shows a portion of the structure of the domain name system. The root node does not have a name;[7] when necessary to represent it explicitly, it is represented with a dot. Below the root are several hundred top level domains or TLDs. There are several types of top level domains. The generic TLDs originally described the kind of organization to which a domain name was assigned.

[7] The root node does not have a name, but the root name servers do. They are a.rootservers.net to m.rootservers.net.

There are about 200 country-code TLDs, such as .us for the United States and .ca for Canada. There is a growing list of TLDs sponsored by organizations, such as .aero for the aviation industry and .museum for museums. There is a small set of reserved names, including .arpa, .local, and .test. The second level domain names, shown as 2LD in Figure 0-10, are registered to organizations or individuals which pay an annual fee to commercial domain registrars. The commercial registrars are, in turn, accredited by ICANN, the Internet Corporation for Assigned Names and Numbers. It is possible to have three or more levels of domain names. Subdomains, if any, are named and managed by the holders of the second level domain names. At the lowest level are names of host computers or network nodes. It is the responsibility of the holders of second level domain names to maintain authoritative name servers. These name servers resolve subdomain and host names within the second level domain. Often the name servers are provided by the domain name registrar, but that isn't required.

Domain names are written with the lowest level on the left, so Kennesaw State University's web server name is www.kennesaw.edu. That is a fully-qualified domain name (FQDN) for host *www* in domain *kennesaw.edu*. It is possible to configure a DNS search suffix list that fills in the domain name, so a client computer on the Kennesaw State campus might be able to reach the web server by typing only www.

The domain name system has some built-in assumptions that the character set is ASCII. Domain names in non-Latin alphabets, such as Cyrillic or Greek are called internationalized domain names. They are implemented by encoding Unicode characters as pure ASCII for internal handling using a scheme called *punycode*.

A client computer, such as a laptop, tablet, or phone, uses very simple software called a stub resolver to query a DNS server on the client's network. If the stub resolver has looked up the name recently, it can return an answer from its cache. If not, it will pass the query to a recursive resolver on its home network. That resolver might also be able to answer from cache. If not, it will pass the query to the next higher level continuing until an answer is received or a root name server is reached. The query then travels down the hierarchy until the authoritative name server for the domain is reached. The authoritative server provides the IP address. Once the client has received an IP address, it is cached to prevent repeated calls to DNS servers.

When a name server is asked for an address for a given name, is can provide an A (address) record with an IPv4 address, an AAAA[8] record with an IPv6 address, or both.

It is possible that a query is sent for a name which does not exist in the domain name system. The proper response from the domain name system is a "no such domain" DNS response. Various ISPs and others have engaged in the unsavory process of presenting advertising instead of the prescribed response.

Although there are only 13 root server names with 13 IP addresses, a technique called **anycast routing** permits a geographically distributed group of over 600 root name servers. Anycast will attempt to send a query to the root server with the fastest path. The IP addresses of the root name servers change very infrequently, and are included in operating system and DNS software releases. If a DNS resolver can reach any root name server, it can find the addresses of all the others.

DNS as originally designed has some serious security shortcomings. DNS Security Extensions (DNSSEC) is an extension to DNS that is beginning to be widely adopted in the third decade of the 20th century. DNSSED uses cryptographic signatures to allow a DNS resolver to verity that the data it received came from the authoritative name server for that domain. It provides integrity protection to allow modification in transit to be detected, and includes a mechanism to guarantee that a "no such domain" response is valid. Adoption has been slow because DNSSEC requires cryptographic signatures from the next higher level in the DNS hierarchy, and those are still typically handled manually.

[8] Because there are four times as many bits in an IPv6 address as an IPv4 address.

IPv4 Addresses and Routing

The 32-bit Internet Protocol address format, called IPv4, was described in an IETF document in the fall of 1981, the same year the original IBM PC was introduced. The choice of a 32-bit address space was made earlier, in 1978 (Huitema, 1996). IPv4 addresses have been in use since 1982 and probably still carry the bulk of Internet traffic.

A 32-bit address space allows for 2^{32} or about 4.29 billion addresses. There are about 600,000 reserved addresses, leaving about 3.7 billion available for assignment. At the time, that was believed to be sufficient for the foreseeable future.

IPv4 addresses are written in dotted-quad form, four numbers separated by periods, for example, 192.168.137.97. Each of the numbers represents eight bits, and so will be in the range zero to 255.

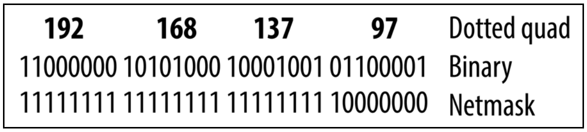
IP addresses consist of two parts, a network part and a host part. The network part identifies a network reachable by routing and the host part identifies one host address within that network. Originally the first eight bits were the network number, allowing for 254 networks. By 1981 the standard was revised so that the first one to four bits of the first octet specified the **network class** that defined the separation between the network part and the host part. You may still hear people talk about class A, B, or C networks. By 1993, the IETF realized that three sizes[9] did not fit all, and introduced classless interdomain routing (CIDR).

CIDR includes the concept of a **netmask**, a bit mask 32 bits long with ones marking the network part and zeros marking the host part. The netmask concept allows division of network part and host part at any bit boundary. Since the netmask is 32 bits, it can be written as a single number indicating the length of the network part of the netmask. The example address above could be written 192.168.137.97/25 with the 25 indicating that the first 25 bits of the address is the network part. In some cases, the netmask is required to be specified in the older dotted quad format. A /25 network would be 255.255.255.128. That's 25

[9] There were actually five classes, with class D being a multicast class, and class E reserved.

bits of ones and seven bits of zeros. The division need not be on a byte boundary, as was required with class-based addresses. The introduction of CIDR did not change the form of the address; it only changed how the address bits are interpreted.

The /25 network would have 32 – 25 = 7 bits of host part, so 2^7 or 128 host addresses. The first address, with a host part of all zeros, is the address of the network itself. The last address, with a host part of all ones, is the network's broadcast address. That leaves 126 addresses available for network hosts or nodes. The number of host bits, *h,* is 32 – *n* where *n* is the netmask width, *i.e.* the number following the slash. The number of available host addresses is 2^h-2 to account for the first and last addresses, the network address and broadcast address, respectively.



192	**168**	**137**	**97**	Dotted quad
11000000	10101000	10001001	01100001	Binary
11111111	11111111	11111111	10000000	Netmask

Figure 0-11
IPv4 address in dotted-quad and binary, showing a /25 netmask in binary

IPv4 reserves a number of addresses for special uses. The **loopback address**, 127.0.0.1, is used for a node to send messages to itself. **Link-local addresses** are in the block 169.254.0.0/16 and are assigned by an operating when neither manual nor automatic IP configuration is available. Link-local addresses were originally intended for automatic configuration. In a managed network, the presence of a link-local address on a machine suggests connectivity or configuration failure.

In addition to the link-local addresses, the IANA has reserved three ranges of IP addresses for private networks: 10.0.0.0 – 10.255.255.255, 172.16.0.0 – 172.31.255.255, and 192.168.0.0 – 192.168.255.255. These private addresses are sometimes called reserved addresses, or RFC 1918 addresses after the number of the standards document that set them aside. They are also sometimes, incorrectly, called non-routable addresses. Since these addresses are reserved for private networks, many different networks can use addresses in the same ranges without conflict. Router / gateway devices for home or small business networks usually assign addresses from one of these ranges automatically by including a dynamic host configuration protocol server. When a network using private addresses must be connected to the Internet, network address translation is required.

Network address translation (NAT) allows a network to use private or link-local addresses internally and provide a connection to the Internet using a single registered address. Registered IP addresses are generally assigned in large blocks to Internet service providers (ISPs) who, in turn, assign them to customers as needed.[10] For home use, most addresses assigned by ISPs are dynamic, meaning they can change from time to time. Blocks of one or more fixed addresses are available, usually at extra cost, for organizations and sophisticated home users.

Network address translation takes advantage of the fact that TCP connections and UDP datagrams are characterized by protocol, the two endpoint IP addresses, and the two endpoint port numbers. A NAT device will have two network interfaces, one for the outside, registered address, and one for the inside, private address. For outgoing packets, the NAT substitutes its registered address for the internal source address and a port number of its own, selected from the dynamic port numbers, for the source port number. The NAT device maintains a mapping of the port numbers it used with internal addresses and port numbers. For incoming packets, the NAT device uses the destination port number to look up the internal IP address and port number, rewrites the destination address and port number, and forwards the packet on its internal interface.

The entries in the NAT device table are made when a host on a private network initiates an outgoing contact. In the absence of special configuration of the NAT device, a host on a private network cannot receive unsolicited network packets. This is sometimes held out as a security advantage.

Port forwarding is necessary to operate a host that can receive unsolicited network packets while within a NAT network. Port forwarding is accomplished by making a semi-permanent entry into the NAT device's translation table. For

[10] In the United States, blocks of IP addresses are assigned directly to very large organizations, to the military, and to some educational institutions.

example, to operate a web server on the machine with address 192.168.137.97 inside a NAT network, the NAT device would be configured to forward all incoming traffic directed to port 80, the HTTP port, to port 80 on the host at address 192.168.137.97. The advertised address of the web server must be that of the NAT device's registered address.

IPv4 Configuration and the Dynamic Host Configuration Protocol

An IPv4 host must, at a minimum, be configured with an IP address, a netmask, and the addresses of one or more DNS servers. If communication off the local network is needed, it must be configured with the IP address of at least one router, the default gateway. The default gateway receives packets for which no other route is configured. For small and medium sized networks, often it is the only router configured. If the host is to be a server and accessible by name, it must have its name to IP address mapping configured in the local DNS servers. The existence of link-local addresses provides one way of configuring the IP address automatically, but netmask, DNS servers, routers, and possible DNS mapping must still be configured. A configuration error can result in an inoperable network host. Some errors, such as duplication of IP addresses, can result in unstable network operation.

Manually configuring the devices for even a small network can take significant effort, especially if the network is to support portable devices that join and leave the network on an *ad hoc* basis.

The solution is to automate all or nearly all configuration tasks with a dynamic host configuration protocol (DHCP) server. Like DNS, DHCP operates at the application layer of the Internet Protocol suite. A DHCP server listens for broadcasts from hosts wanting to join the network. The broadcast for a DHCP server is sent on the local link network, usually Ethernet, so an IP address is not needed to send the broadcast nor receive the response. When a DHCP broadcast is received, the server will supply configuration information to the host client, which will use the configuration information to complete the configuration.

DHCP servers can configure all of the required parameters and many others, including things like the address of a network time protocol (NTP) server, which the client can then use to obtain the correct time. Some DHCP servers integrate with DNS to provide the necessary DNS records when an IP address is assigned.

DHCP servers assign IPv4 addresses from a pool of addresses specified when the DHCP server is configured. Since these routers generally also include NAT, common practice is to configure a pool of addresses from the RFC 1918 allocations, often 192.168.0.0/24 or 10.0.0.0/24.

Most routers designed for home or small office use incorporate a DHCP server so that client computers are configured automatically.

IPv4 Address Exhaustion

By 1992, people planning what was then called IPng[11] projected that the 32-bit address space would be exhausted between 2005 and 2015 (Huitema, 1996). The regional Internet address registries did run out of IP addresses near the end of the predicted time. IP addresses are assigned to regional registries by the Internet Assigned Number Authority, which ran out of addresses in 2011. The regional registries still had addresses to allocate. The American Registry for Internet Numbers exhausted its supply of addresses in 2017, and the registry for Europe, Middle East and Central Asia exhausted its supply in 2019. ISPs still have pools of IPv4 addresses to allocate to customers in 2020. A major crisis was avoided through using CIDR to allocate addresses in blocks smaller than 256 and using NAT to provide internal addresses for many networks. The regional registries can also reclaim addresses not in use.

The long-term solution is an Internet protocol with larger addresses, now called IPv6. The mechanism that was adopted includes 128-bit addresses, offering the potential of 2^{128} IPv6 addresses. To put that in perspective, the mass of Earth in grams is about 2^{92}. There are enough IPv6 addresses to assign one to every gram of Earth's mass with 2^{36} extra. (As a practical matter, not all 2^{128} addresses are available for assignment because some are reserved, but there are more than enough.)

[11] IP new generation.

IPv6 Addresses and Routing

IPv6 addresses are written as eight groups of four hexadecimal digits separated by colons like this: fe80:0000:0000:0000:e2b7:00ff:0000:ada0. That's a lot to write, but many IPv6 addresses can be shortened. Leading zeros in any group are not significant and can be omitted. So, 00ff above could be written as ff only. Two or more consecutive blocks of all zero can be shortened to two consecutive colons, but only once. If there are more than two such blocks, only one can be shortened because the length of the shortened block is computed from the length of the block as written. Shortened, the address above is written as fe80::e2b7:ff:0:ada0.

IPv6 addresses have prefixes that serve the purpose of the netmask in IPv4 addresses, namely to mark the portion of the address used for routing packets. The notation is the same as the CIDR notation, a slash and number following the address. For IPv6, the number must be from one to 128. The default prefix is /64. That is, the leftmost 64 bits are used for routing.

Each IPv6 address also has a **reachability scope**. The **node-local** address is ::1/128, or all zero followed by a single one. The node-local address is equivalent to the IPv4 loopback address, and is used when a host wants to send a message to itself. Messages sent to the node-local address are never sent to a network interface controller.

Addresses with **link-local** reachability are used for sending datagrams to hosts on the same local network. Unlike IPv4 link-local addresses, which are a kind of last resort, every IPv6 node has a link-local address that is configured automatically. Link-local addresses have the form fe80::/10. The example address at the beginning of this section is a link-local address. The link-local reachability scope also includes a link-local multicast address and the unspecified address. The unspecified address is all zero, and is used before a host or node has received an IPv6 address. The default route is also all zero, but with a different prefix: ::/0.

Addresses with **global** reachability are called aggregatable global unicast addresses and are equivalent to public IPv4 addresses. Such addresses are called aggregatable because, with proper assignment, it is possible to arrange IPv6 addresses in a hierarchy, where the next router is “up” the address hierarchy until a datagram reaches the point when it must leave that portion of the network. It is also possible to route IPv6 packets in much the same way as IPv4 packets.

Multicast addresses also have global reachability and are intended for one to many communication. They have the form ff00::/8.

IPv6 is designed for full **stateless autoconfiguration**. When a device joins an IPv6 network, it generates a link-local address. The first 64 bits start with fe80. The last 64 bits are formed from the device’s MAC address. Once a device has a link-local address, it sends a router solicitation (RS) request to the default multicast address. A router will respond with a router advertisement (RA) response which includes the router’s address and also the addresses of DNS servers (Dooley, 2015).

Once a device has a router address, it can generate a unique global address from the first 64 bits of the router address and the last 64 bits from the device’s link-local address.

Devices also need to be able to find the MAC addresses of other hosts on the local network. With IPv4, that was done with ARP. The equivalent for IPv6 is the neighbor solicitation protocol. The device sends a neighbor solicitation packet, and the host with the address in the NS packet responds with a neighbor advertisement packet, which includes the MAC address. (Dooley, 2015)

Network hosts or nodes may need information not provided through autoconfiguration. That can be provided with dynamic host configuration protocol version 6 (DHCPv6). Routers can also be configured through DHCPv6.

IPv6 Transition

Although IPv6 solves the address exhaustion problem, IPv6 and IPv4 addresses are not interoperable. There have been a number of approaches to the problem of making the transition to IPv6 that also accommodate both clients and servers that still use IPv4. The accepted solution at the end of 2020 is dual-stack IP on clients and network infrastructure such as routers.

Both IPv4 and IPv6 can coexist “on the wire.” The unit of transmission on a typical local area network is the Ethernet frame, and the IP datagram is payload, never used as other than data by the underlying network. The problem is to allow clients to process both IPv4 and IPv6 packets. The preferred solution as of the end of 2020 is the dual stack environment, as shown in Figure 0-12. The dual stack environment is supported by current versions of all major operating systems.

The distinction between IPv4 and IPv6 is the version number in the datagram header. Dual-stack software uses the version number to select the appropriate datagram handling code. Once Internet Protocol layer processing is complete, the packet or datagram can be passed to the transport layer.

Microsoft’s implementation, beginning with Windows Vista, is typical. Dual stack applications should expect to process only IPv6 packets or datagrams. If an IPv4 packet is received, it is converted to IPv6 format by using the IPv4 address as the low-order 32 bits of an IPv6 address and the constant 0:0:0:0:0:ffff as the high-order 96 bits.[12] When the application is sending, lower protocol layers recognize the prefix and select the IPv4 branch of the dual stack.

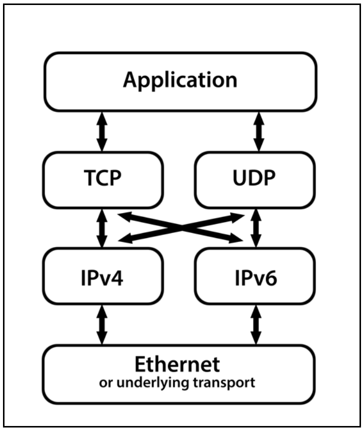


Figure 0-12
Dual stack IPv4 / IPv6 environment

0.5 Physical Transmission

Up to this point, the chapter has focused on the logical interpretation of bits and given little attention to the mechanism by which bits are transported from one network node to another. Physical transmission is important because the choices made affect the distance over which signals can be transmitted and the maximum speed of transmission.

[12] The use of 0:0:0:0:0:ffff is an IETF standard, described in RFC 4291, and not a Microsoft extension.

0.5.1 Network Topologies

The topology of a network describes how network signals travel, and is easiest to think about in terms of wired networks although wireless networks also have topology. The physical topology of a network is the layout of the wires or other media that carry the signals. The logical topology describes the flow of the signals themselves. Logical and physical topologies are often the same, but do not have to be.

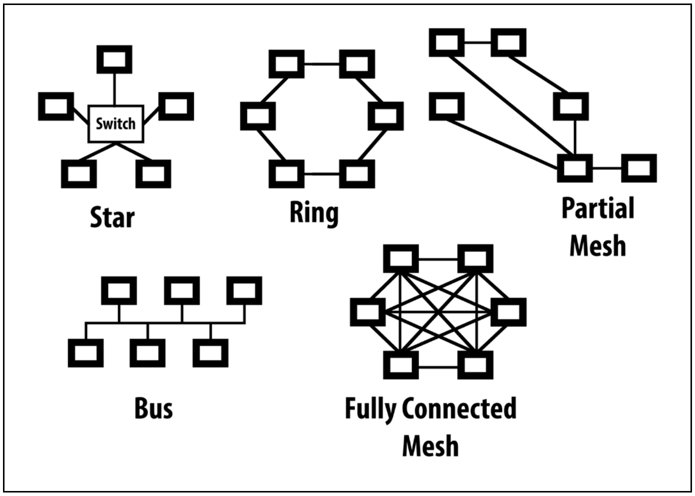


Figure 0-13
Some physical network topologies

Figure 0-13 shows some physical topologies. Modern wired networks are most often physical and logical star networks. As discussed in Section 0.3.2, modern Ethernet networks connect each host or node to a central switch that is responsible for forwarding incoming Ethernet frames to the proper destination. Connections between the network nodes and the switch are **point-to-point connections**.

The original Ethernet design was a bus topology. Every node on the bus received every transmission and network interface controllers were responsible for processing broadcast messages and those with the controller's own address and ignoring others. A physical bus as shown in Figure 0-13 is a **multipoint connection**. A physical bus is difficult to install and maintain, especially when adding new network nodes. Ethernet hubs, which just connected all ports together, made a physical star and logical bus network.

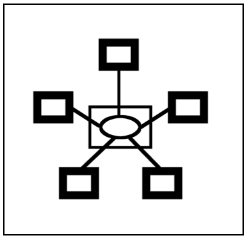


Figure 0-14
Physical star, logical ring

IBM's token ring network was at first wired as a physical ring. Every node received messages and passed on those not addressed to it. Like a physical bus, a physical ring is difficult to install and maintain. IBM's introduction of the medium access unit allowed a logical bus network to operate in a physical star configuration, as shown in Figure 0-14.

A partial mesh network requires more complex software at each node than network architectures where each node is responsible only for processing its own and broadcast messages. In a partial mesh, some nodes must be able to forward messages for other nodes. A sparse network is one in which the number of links is much less than the maximum possible number.

A fully-connected mesh provides the greatest redundancy, but its complexity, approximately $n^2 / 2$ where n is the number of nodes, increases very rapidly.

Particularly in larger organizations, networks may be arranged in a hierarchical structure. Figure 0-15 shows such an arrangement. The diagram shows three star networks, possibly for organizational units, each potentially with their own servers. More or fewer connected networks are possible. Each star network switch is connected to a core switch that provides connectivity to common resources such as enterprise storage and a router to external networks.

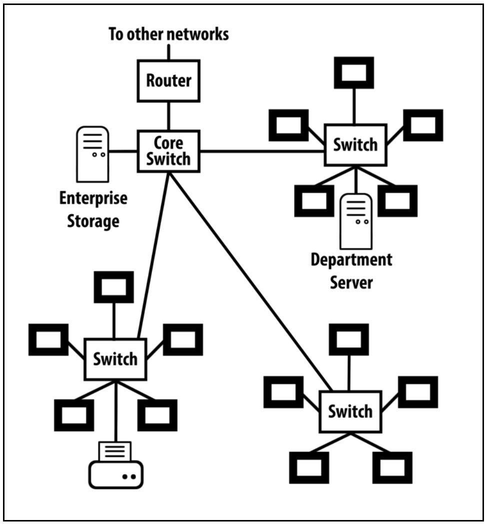


Figure 0-15
A hierarchical networks

An arrangement like that of Figure 0-15 can simplify wiring, especially if the nodes in each of the star networks are physically close together. However, connecting several networks increases the size of the broadcast domain [13] . Broadcast messages will be sent to every node on the network. When a new device enters the network, messages to it are flooded to every port other than that on which the message entered the switch until the switches learn the location of the device. That is described in Section 0.3.2. While not strictly broadcast messages, flooding has a similar effect on network bandwidth.

One solution to limiting the broadcast domain is the **virtual local area network**, or VLAN. The VLAN standard in widest use is IEEE 802.1Q which identifies virtual LANs with a 12-bit tag field included in the Ethernet frame. The first and last values are reserved, making the number of possible VLANs on a single physical network 4,094. Other IEEE 802.1 standards offer up to 16 million VLANs on one network. Prior to adoption of IEEE 802.1Q, several manufacturers of communication equipment implemented VLANs using proprietary techniques.

In addition to minimizing broadcast domains, properly configured VLANs can improve network security. If the departmental server in Figure 0-15 held confidential information, the traditional approach to protecting it would be using passwords, perhaps with two factor authentication.[14] A properly-configured

[13] Do not confuse broadcast domain with collision domain. They are different. See Section 0.3.2.

[14] Two factor authentication is discussed in Chapter 7. ****

VLAN can limit which workstations are able to access the server, thereby reducing the attack surface.

Software defined networks expand on the idea of VLANs and include the ability to automate and optimize network configurations. Software defined networks are most suitable for large enterprises.

0.5.2 Analog and Digital Signaling

Data transmission depends on a signal and a medium to carry the signal. Information is transferred through changes in the signal as a function of time. A communication channel is characterized by:

- The medium used
- The signal transmission method
- Direction(s) in which a signal can flow
- Susceptibility to attenuation, noise, and distortion
- Bandwidth
- Time delay and time jitter.

There are two classes of transmission media, guided and unguided. In **guided media**, also called bounded media, signals are guided along a path. Copper wire and fiber-optic cable are examples of guided media. **Unguided** or unbounded media use electromagnetic waves to transmit data. Examples are radio, microwave, and, for short distances, infrared light.

Signal transmission can be analog, that is, continuously varying within a range of values, or discrete. In discrete transmission, the signal takes on a countable set of values.

A transmission medium that can carry signals in both directions simultaneously is called **full duplex**. A medium that can carry signals in either direction, but only one at a time is called **half duplex**. A medium that carries signals in one direction only is called **simplex**.

Attenuation is the loss of signal strength, primarily due to distance, but other causes are also possible. For example, a wall will attenuate a Wi-Fi signal. Signal wires can act as antennas and pick up unrelated electromagnetic radiation, called **noise**. Noise from external sources is called interference. Noise is measured at the end of a channel distant from the transmitter as the ratio of the power of the signal to the power of the noise component. That ratio is called the **signal to noise ratio**. If signal and noise are expressed on a logarithmic scale, the resulting ratio is in decibels. **Distortion** changes the shape of the waveform, particularly when signals are transmitted over long distances.

The **bandwidth** of a transmission medium is the range of frequencies it can carry without excessive attenuation. Because the bandwidth of a transmission medium limits the number of bits per second the medium can carry, the word "bandwidth" is often used to express the data-carrying capacity of a medium in bits per second.

Latency is the time from transmission of a signal to its receipt, and is unavoidable. Although electrical and electromagnetic signals travel at a large fraction of the speed of light, distance and processing at intermediate nodes can cause perceptible delay. Grace Murray Hopper used to tell the story of the Navy admiral who asked her why communication via satellite took so long. Dr. Hopper answered, "Because there is a very large number of nanoseconds between here and the satellite."[15] **Jitter** is variation in latency, and can be more troublesome than delay, particularly for applications like voice and video.

[15] See the illustration of "nanosecond" in Figure ****.

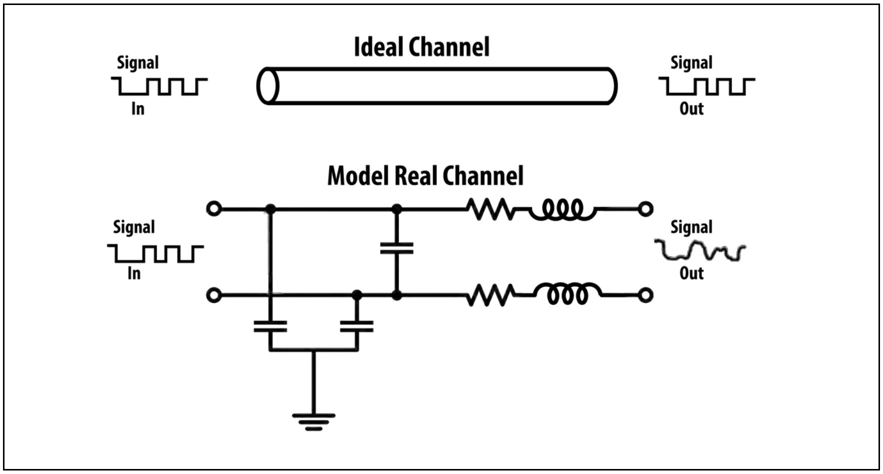


Figure 0-16
Ideal and real communication channels

Figure 0-16 compares an ideal communication channel, represented as a pipe, with a model representing an actual communication channel using copper wiring. In an ideal communication channel, a signal would go in one end and come out the other unchanged after an invariant delay proportional to the channel length.

In a real channel, the signal is changed by attenuation, noise, and distortion of the wave form. Capacitance (parallel line symbols) changes the waveform and can admit noise from external sources. Resistance (saw tooth symbol) attenuates the signal. The striped triangle is a ground symbol, showing that some of the signal strength can be drained away during transmission. Inductance (coil symbol) distorts the signal and can admit external noise. A real channel does not actually contain the capacitors, resistors, or inductors in the diagram. Instead, the conductors themselves exhibit these characteristics, which increase with the length of the channel.

In digital transmission, a signal such as the one shown at the output of the model real channel can be recovered exactly provided it has not been too degraded by attenuation, noise, or distortion. This is possible because it is a discrete signal. The expected number of levels of the signal is small, from two in a pure binary signal, up to about five with other kinds of discrete encoding.

Baseband and Broadband Circuits

The term “broadband” has been misused to mean “fast.” In discussions of communication technology, more precision is needed. A baseband circuit is one in which the signal covers the entire frequency spectrum from zero Hz up to the maximum frequency the channel can carry with acceptably low attenuation. A gigabit Ethernet circuit, often held out as the goal to achieve in home Internet connections, is a baseband circuit. In fact, “base” in 1000BASE-T means baseband.

A broadband circuit is one in which two or more frequency bands carry two or more independent signals. A digital subscriber line (DSL) Internet connection is a broadband connection. The frequency band from 300 HZ to 3,400 Hz is used to carry voice traffic. Two or more bands are allocated beginning at 4,000 Hz. At least one band carries data from the customer’s equipment to the carrier, and at least one band carries data from the carrier to the customer. If the distance to the carrier’s equipment is short and the copper line is in good condition, many more bands can be used, increasing the speed of the DSL connection.

Television and radio are broadband technologies, with each television channel or radio station occupying a frequency band within some allocated spectrum, for example 89.5 MHz to 108 MHz for FM radio in the United States. The technique of carrying two or more signals within some frequency band is called **frequency-division multiplexing**.[16]

Baseband Signaling Mechanisms

The stream of bits that comprises a digital signal looks like the diagram labeled “Input Signal” in Figure 0-16. Such a signal is called non-return to zero because there is no space or rest between each bit. Recall from Section 0.3.1 that the preamble of an Ethernet frame is 56 bytes of alternation zero and one bits, followed by a start of frame delimiter. The preamble allows the receiver to adjust timing with the transmitter so that both agree on where the bit boundaries are.

[16] Packet switching, in which packets from multiple sources are sent one after another on a transmission medium, is one form of **time-division multiplexing**.

The receiver then samples the incoming signal in the middle of each bit-time. The preamble establishes bit boundaries. The start of frame delimiter establishes where each byte of the incoming signal begins.

Although the receiver is synchronized with the transmitter at the beginning of each frame, a long string of either ones or zeros would cause no changes in the signal during that time. It is possible for the receiver's timing to drift slightly during a period of no transitions. Especially at high speeds, that can mean either a missed bit or an erroneous bit inserted into the signal stream. Everything after such an error will be incorrect.

To prevent such errors, baseband transmissions are encoded in such a way that there are regular transitions in the signal even when long strings of ones or zeros are transmitted. Ten megabit Ethernet uses Manchester encoding. According to the IEEE 802.3 standard,[17] a zero bit is represented by a high-to-low transition in the middle of the bit period, and a one bit is represented by a low to high transition. That means there will always be a transition in the middle of each bit time, and often there will be two, one at the beginning and one in the middle. For example, if two ones are sent in sequence, there must be a high to low transition at the end of the first bit period so that there can be a low to high transition to represent that second one. With at least one transition per bit time guaranteed, the transmitter and receiver clocks can be kept in synchronization. However, since there will frequently be two transitions per bit time, the bandwidth of the medium must be at least twice the data rate. In fact, Manchester encoding requires that the transmitter have a clock that operates at twice the bit rate of transmission as shown in Figure 0-17. A close look at the figure will show you that the Manchester encoded signal is the XOR of the clock and data signals.

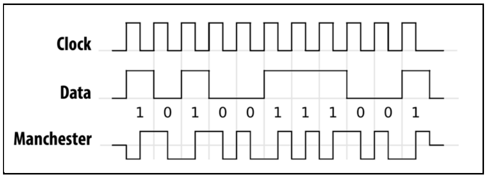


Figure 0-17
Manchester encoding (IEEE 802.3)
Stefan Schmidt
Wikimedia Commons

[17] In the original Manchester encoding, invented by G. E. Thomas at the University of Manchester about 1949, the transitions are the other way around.

The requirement that the transmission medium have a bandwidth of twice the data rate makes Manchester encoding unsuitable for higher data rates. Hundred megabit Ethernet encodes each four bits into a five-bit group in such a way that there will be at least one transition for each five-bit group. That is called a 4B5B code. The five-bit groups are then transmitted using a three-level code called MLT-3 for copper connections. It's still a discrete signal, but the values are a positive voltage, zero, and a negative voltage. For fiber optic transmission, the 4B5B code is encoded using a method called non-return to zero inverted, or NRZI in which one and zero bits are sent as transition or no transition, rather than as levels. The 4B5B encoding means there will always be at least one transition per group to keep transmitting and receiving clocks synchronized.

Successively higher speeds use different encoding mechanisms to be sure signal transitions occur frequently enough to allow the receiver's clock to remain synchronized, and to limit the bandwidth required from the transmission medium.

Broadband Signaling Mechanisms

A broadband connection is one in which two or more frequency bands carry independent signals. That means there must be two or more carrier signals of different frequencies. Information is transmitted on the carrier signals by modulating, that is changing, the carrier signal in a way that represents the transmitted data stream. The received data stream is recovered by demodulating the carrier signal. This is performed by a device called a modulator/demodulator, or **modem**.

The fundamental waveform used in broadband signaling is a sine wave. Sine waves have the properties of frequency, amplitude, and phase. One or more of these properties are modulated, *i.e.* changed with respect to time, to transmit data. Frequency is the number of repetitions per second. Amplitude is the power of the wave. Power measurements for electromagnetic waves are complex. What is important in signaling is that the received power be sufficient to have a signal to noise ratio greater than one at the lowest amplitude used.

For radio waves, frequency is measured in Hertz. For light waves, frequency is measured in nanometers (nm) between peaks of the waveform. A typical wavelength for multimode glass fiber is 850 nm, which is 352,697 GHz.

The phase of an electromagnetic wave is expressed in degrees from some reference. The upward zero-crossing of a sine wave is zero degrees, the peak is 90 degrees, and the downward zero-crossing is 270 degrees. Often phase angle is measured with respect to the carrier wave before a phase change so that it isn't necessary for the receiver to maintain a reference waveform.

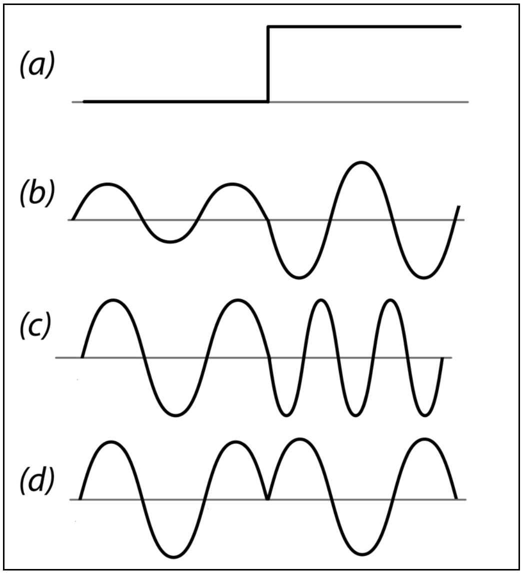


Figure 0-18
Modulation technique; (a) digital signal; (b) amplitude modulation; (c) frequency modulation (d) phase modulation

Figure 0-18 illustrates modulation of the three properties. Figure 0-18 (a) shows the transition of a digital signal from low to high. Figure 0-18(b) shows amplitude modulation of a sine wave for such a transition. Figure 0-18(d) shows frequency modulation, and Figure 0-18(d) shows phase modulation.

The unit of modulation rate is the **baud**,[18] equivalent to one signal event, *i.e.* one change in the carrier signal, per second. Modulation can encode more than one information bit in one signal event. For example, quadrature amplitude modulation (QAM) as used in Wi-Fi and some fiber optic systems modulates both phase and amplitude. With QAM, four or more bits are transmitted with each change in the carrier.

[18] Named after Émile Baudot, 19th century French telegraph engineer, who invented a telegraph code and a mechanism for time division multiplexing of telegraph messages.

Demodulating a broadband signal produces a stream of bits. As with baseband signals, there must be a way to maintain synchronization of receiver to transmitter. So, data transmitted on broadband systems are encoded using Manchester coding, 4B5B, or a similar mechanism to be sure bit transitions occur frequently enough to maintain synchronization.

Transmission Media

Guided transmission media, copper wire or fiber optic cable, is used in permanent or semi-permanent installations. Unguided media, radio waves and sometimes infrared, are used for portable devices or where installation of guided media is not practical, such as in an apartment building. Focused unguided media are also used for long distance communication, such as with satellites.

The principle types of copper transmission media are coaxial cable and twisted pair cable. Both coaxial cable and twisted pair cables can have either solid conductors or stranded conductors. Solid conductors are used for permanent installation. Stranded conductors, which withstand flexing better, are used for applications like patch cables.

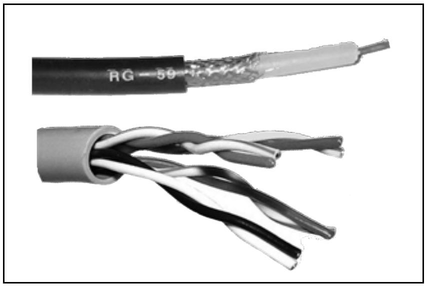


Figure 0-19
Coaxial cable (top)
twisted pair (bottom)

Coaxial cable has an insulated inner conductor covered by a conductive braid. The braid is covered by an insulating jacket. Conductor, insulator, braid, and jacket have a common center or axis, hence the name coaxial. Coaxial cable is used in broadband circuits with multiple channels, such as television or other audiovisual systems because of its high bandwidth.

Twisted pair cable is used for baseband transmission at speeds up to 10 GHz. Twisting the pairs means that interference affects both wires in a pair approximately equally. Twisted pair cables for Ethernet have four pairs of wires, eight conductors total. The pairs are color-coded for matching at the ends of the cable. Different pairs in a cable have different numbers of twists per meter, called pitch, to minimize interference from within the cable, known as **crosstalk**. Cable intended for higher speeds will have more twists per meter. It may have an overall foil shield or each pair may be separately shielded. There may be a bare drain wire that is in electrical contact with the shield material and that is intended to be grounded. The drain wire, if present, should be grounded only at one end of a cable run.

Specifications for twisted pair cable are established by the Telecommunications Industry Association (TIA)[19] and the ISO / IEC. Specification for twisted pair cable are given as categories, with Category 5e being used for Ethernet up to one gigabit per second. Categories 6 and 6A are compatible with category 5e, but with higher bandwidth specifications. Categories 7, 7A, and 8 are defined for specialized uses.

Fiber optic cables transmit light waves rather than electrical signals. The high frequency of light waves make optical fiber transmission suitable for very high bandwidth applications. Optical fiber transmission is not affected by external electromagnetic noise. Construction of a fiber optic cable is shown in Figure 0-20. The core is a glass fiber roughly the thickness of a human hair. The core is surrounded by glass or plastic cladding with a refractive index less than that of the core. Light hitting the boundary between core and cladding is reflected back into the core rather than being allowed to escape. The process of total internal reflection means light travels along the core. The cladding is surrounded by a protective coating, then a strength member, often aramid fibers,[20] then a protective jacket. It is common to include a number of fibers with cladding and coating within one strength member and jacket.

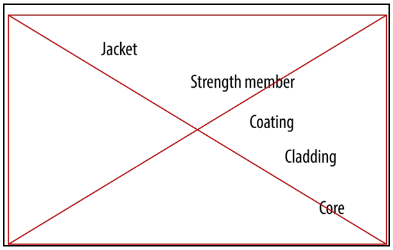


Figure 0-20
Structure of fiber optic cable
OFS Fitel

Near infrared light is injected into fiber optic cables with lasers or LEDs and detected at the receiver with a photodetector device. The light signal is made to

[19] TIA was formerly part of the Electronics Industries Alliance, which ceased operations in 2011. You may still see “EIA” specifications for networking equipment and media.

[20] One brand of aramid fiber is Kevlar, used for making bullet-resistant vests.

carry data by modulating the light signal. Wave length division multiplexing allows signals of two or more different wavelengths of light to be carried in the same fiber.

Fibers for fiber optic transmission are either multi-mode or single mode. Multimode fibers have a larger core, which makes connected equipment less expensive. However, the larger fiber size reduces the bandwidth and also the distance over which signals can be transmitted. Single mode fiber is much narrower, requiring more precise connectors and electronics. It is suitable for longer distances. It is important that the type of fiber match the equipment to which it will be connected.

Unguided communications use radio waves, or for very short distances, infrared light. In the United States, frequency allocations in the radio and microwave spectrum are controlled by the Federal Communications Commission (FCC). Operation of transmitting devices over most of the radio and microwave spectrum requires a license form the FCC. However, the FCC has set aside some frequency bands for unlicensed operation, and allows unlicensed operation in some other frequency bands at very low power. In the United States, unlicensed does not mean unregulated. The FCC has stringent requirements for power and frequency even for unlicensed uses.

Unlike fiber optic transmission and some wired transmission, radio frequency bands are divided into channels, and it is the bandwidth of the channel that is available for modulation. For example, some Wi-Fi devices operate in the 2.4 GHz band, which is divided into eleven overlapping channels of 22 MHz each. It is the 22 MHz bandwidth of a channel that is available for modulation. Wi-Fi and Bluetooth use spread-spectrum technology to minimize interference.

In the United States, wireless local area networking (**Wi-Fi**) operates mainly in the 2.4 GHz and 5.0 to 5.9 GHz bands. Although the higher frequencies allow potentially higher throughput, they are also more susceptible to blocking by walls and other obstructions. In 2020, the FCC allocated space in the 6 GHz band for wireless local area networks. The Wi-Fi Alliance has defined “generations” of Wi-Fi, with the most current being Wi-Fi 6, which is equivalent to the IEEE 802.11ax standard. When 6 GHz equipment is available, it will be branded Wi-Fi 6E.

Wi-Fi is generally operated in infrastructure mode, with one or more access points providing connectivity to portable and mobile devices. It is also possible to operate Wi-Fi devices as peer-to-peer networks without access points. With specialized antennas, Wi-Fi can be used for point-to-point transmission over line-of-sight distances of hundreds of meters.

The design of Wi-Fi attempts to secure it from unauthorized access or interception. The first attempt, wired equivalent privacy (WEP) had a major flaw in the encryption algorithm. WEP was replaced by Wi-Fi Protected Access (WPA), which also had flaws. The current version is WPA3, which supersedes WPA2.

Bluetooth was originally intended as a replacement for cables that connect devices to personal computers. It is usually lower-powered and shorter range than Wi-Fi, although a maximum range is not specified. Bluetooth also operates in the 2.4 GHz band. It is now mostly used to connecting devices to mobile equipment, such as a headset to a cellular telephone, although keyboards and mice with Bluetooth connections are common. Bluetooth 3.0 can provide speeds up to 3 Mb/second at a cost of increased power consumption. Bluetooth Low Energy minimizes power consumption through reduced speed and a much-simplified protocol stack.

There are other wireless networking technologies, for example ZigBee for low power and low speed personal area networks and LoRa for low power longer range networks.

Cellular communication is provided by a number of competing cellular carriers that hold the necessary FCC licenses and operate the infrastructure. Subscribers' cellular devices are covered by the carrier's FCC license. Until about 2020, there were two technologies, code division multiple access (CDMA) and the Global System for Mobile Communications (GSM), used by different carriers. The GSM standard is maintained by the European Telecommunications Standards Institute (ETSI) and legally mandated in the European Union. ETSI establishes "generations" of mobile communication standards. United States carriers are completing a transition to 4G LTE,[21] a step in the direction of full 4G compliance. When the transition is complete, early in the 2020s, there will be no public CDMA networks in the United States.

Cellular devices are increasingly used to transfer data rather than making voice calls. Many cellular devices can function as Wi-Fi “hotspots,” allowing Wi-Fi (802.11) equipment to make a data connection over the cellular network.

Cellular data speeds are increasing. At the time this is written, carriers are implementing the 5G (fifth generation) cellular standard, which will provide gigabit speeds. Doing so requires higher-frequency signals that do not penetrate obstacles well. That means slow adoption of 5G or greatly expanded infrastructure, or both.

0.6 Summary of Learning Objectives

> *This section of the chapter tells you the things you should know about, but not **what** you should know about them. To test your knowledge, discuss the following concepts and topics with a study partner or in writing, ideally from memory.*

Devices that are always connected to each other and the global Internet typify the 21st century. Networks have many parts that must work together smoothly. Modern networking involves the connection of many local area networks.

Networking consists of messages, protocols, and channels. Modern networks are packet switched, and at the lowest levels, connectionless. Protocols may be reliable or unreliable.

The OSI model is an abstract model of a protocol stack. The Internet Protocol model is an implementation of a protocol stack.

The predominant local area networking technology is Ethernet, which may be wired or wireless. Ethernet network interface controllers have physical addresses distinct from a device’s logical address. The unit of transmission is the

[21] Long Term Evolution.

frame. Ethernet was originally a bus topology, which allowed for collisions and required a mechanism for recovering from them. Switched Ethernet can eliminate collisions, but introduces the possibility of broadcast storms. Broadcast storms can be addressed by algorithms implemented in the switching system.

The Internet Protocol is the principal protocol for internetworking. It implements protocols for both reliable and unreliable communication. Other parts of the protocol specification address security and access specifically to web pages.

There are two incompatible formats for Internet Protocol addresses with very different sizes, structures, and facilities. One way to handle the incompatibility is dual-stack software. The domain name system is used to map structured domain names to logical addresses. The initial standard address format, IPv4, is running out of available addresses. Network address translation allows blocks of reserved addresses to be reused among many private networks.

An Internet Protocol host or node needs several configuration options. IPv4 hosts can be configured manually or using the dynamic host configuration protocol. The IPv6 protocol was designed for fully automatic host configuration. An IPv6 host will have several addresses with different reachability scopes, including node-local, link-local, and global unicast addresses.

The physical transmission medium provides the mechanism for moving bits from one host or node to another. Physical media can be guided or unguided. Several network topologies are available, including star and hierarchical topologies. Physical and logical topologies can be different.

Signaling involves changing the state of the transmission medium over time. The signal can be analog or digital. Transmission can be in one or both directions. Transmission is subject to latency, jitter, attenuation, noise, and distortion. The signal to noise ratio is an important measure of the quality of transmission.

The bandwidth of a transmission medium limits the maximum rate at which information can be transmitted. Transmission circuits are categorized as baseband or broadband, depending on the number of independent signals the channel can carry. Signaling mechanisms differ between baseband and broadband media. In either case, there must be a way to synchronize the transmitter and receiver clocks,

Physical transmission media are principally copper wire or fiber optic cable. There are two major types of copper wire with different characteristics and uses. There are also two major types of fiber optic cable.

Unguided transmission over more than very short distances uses radio waves. In the United States, most local area wireless networking uses frequency bands that do not require licenses for transmitters. Unguided communication of voice and data over longer distances is provided by commercial cellular operators who hold licenses for the frequencies they use. Cellular communication standards are names with generation numbers. The most recent, not yet widely implemented, is 5G.

0.7 References

Dooley, K. (2015, May 12). *What Every Network Admin Should Know About IPv6.* Retrieved from Auvik: https://www.auvik.com/franklyit/blog/ipv6-network-design/ on September 6, 2020

Huitema, C. (1996). *IPv6: The New Internet Protocol.* Upper Saffle River, NJ: Prentice-Hall.

Metcalfe, R. M., & Boggs, D. R. (1976, July). Ethernet: Distributed Packet Switching for Local Computer Networks. *Communications of the ACM, 19*(7), 395-404. doi:10.1145/360248.360253