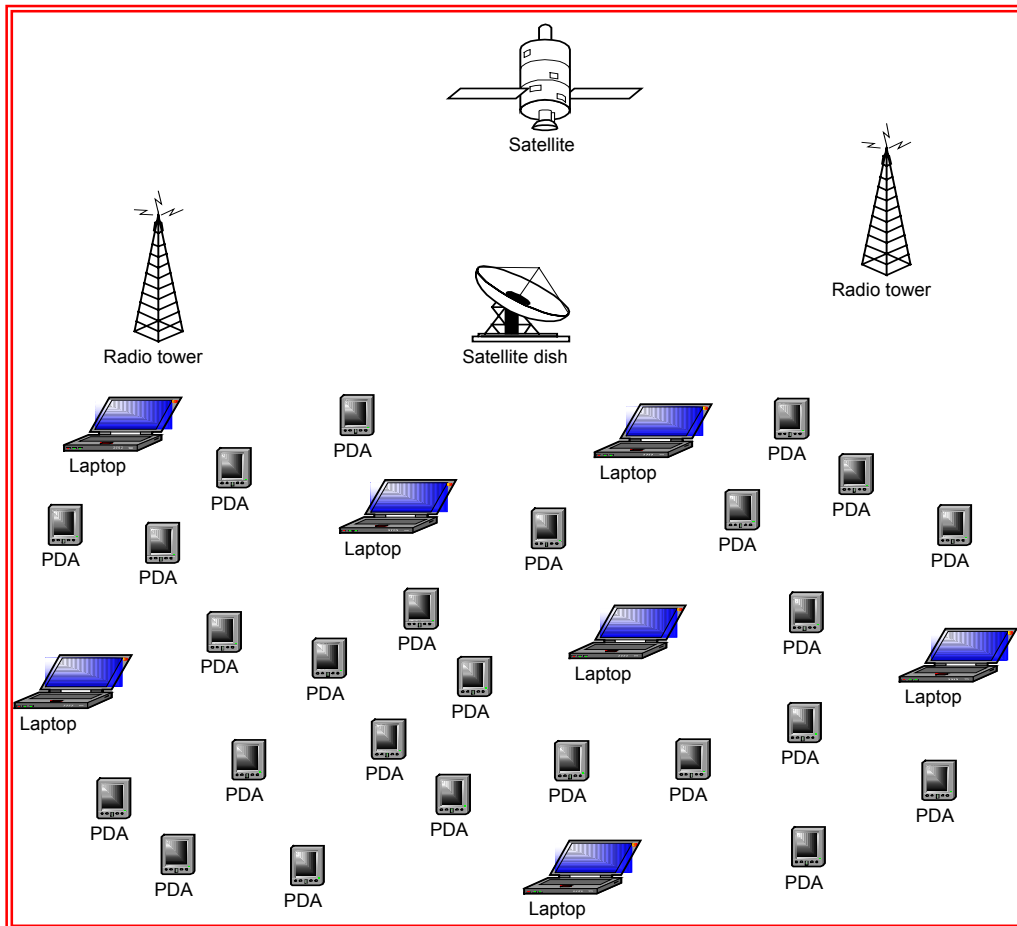


MOBILE COMPUTING

WIRED NETWORK BASICS



Vijay Kumar
Computer Science
University of Missouri-Kansas City
Kansas City, MO 64110, USA.
kumar@cstp.umkc.edu

References

1. Wireless Communications; Principles and Practice by Theodore S. Rappaport. Prentice Hall, 1996.
2. Internetwork Mobility. The CDPD approach, by Mark S. Taylor, William Waung, and Moshen Banan. Prentice Hall, PTR, 1997.
3. Mobile Telecommunications Networking by Michael Gallagher and Randall Snyder. McGraw-Hill, 1997.
4. Wireless and Mobile Network Architectures, by Yi-Bing Lin and Imrich Chlamtac. John Wiley, 2001.

Computer Network

A computer network may be described as a collection of computers and terminals connected together by a wired or wireless medium. A wired medium connects two or more components through hard wire (copper, optical fiber, etc.) and a wireless medium establishes communication through waves such as radio, microwaves, etc. With reference to this simple definition we define the following terms:

Node: A processing unit (computer), capable of performing desired computation independently. A node is also referred to as a **site**.

Protocol: A set of rules governing some operation that is likely to be performed by more than one node. For example, locking protocol, communication protocols etc. The set of rules defines how an operation can be performed among different interacting agents (nodes).

Interface: Defines the means by which one component makes use of the services provided by another component (may not be a node), and the rules and formats for exchanging information across the boundary between adjacent components within a single station. The interface may be specified in terms of its mechanical, electrical, timing or software characteristics, i.e., the interface may be physical or logical.

Geography: Interconnected nodes may be confined to a room or may be spread across the world, and can be classified on the basis of their spatial (geographical) separations.

The geographical distance among the connected components, the bandwidth limitations, the mode of communication, are some of the essential parameters, which identified the type of wired network such as Local Area Network (LAN), Metropolitan Area Network (MAN), and Wide Area Network (WAN). The key issue that distinguishes WAN technologies from LAN technologies is *scalability*. The scalability property defines the capability to grow in every aspects, that is, a WAN must be able to grow as needed to connect many sites (nodes) in a large geographical space. Also from communication viewpoint a WAN should be capable to handle communication among a large number of computers and can grow with demand.

LAN

LANs are usually privately owned and generally confined to a single building or a campus. Most computer science departments' computers, such as PCs, workstations, etc., are connected through LAN. A LAN is usually identified by (a) its size, (b) transmission technology, and (c) the network topology. LANs are restricted in size, which means that the worst-case transmission time is bounded and known in advance. LANs often use a transmission technology where all computers are connected to a single cable, run at speeds ranging from 10 to 100Mbps, and generally make very few transmission errors.

MAN

It is basically a bigger version of LAN and normally uses a similar technology. It spans to a larger geographical area and can support more number of computers (nodes). A MAN also has one or two cables for establishing communication among its nodes. A key aspect of a MAN is that there is a broadcast medium to which all the computers are attached. The main reason for

distinguishing MANs as a special category is that a standard has been adopted for them, and this standard called Distributed Queue Dual Bus (DQDB - 802.6 IEEE Standard) is now being implemented. DQDB consists of two unidirectional buses (cables) to which all the computers are connected.

WAN

A WAN spans a larger geographical area, which may include a country or a continent. It also connects a large number of computers (hosts or nodes) and the connecting network is usually called a communication *subnet* or just a *subnet*. In a WAN the connected components (computers) are of two categories (a) transmission lines and (b) switching elements. The transmission lines move data bits from one computer to another and the switching elements manages the movements of these data bits. The switching elements are special computers, which accepts the incoming messages and forwards them towards the destination computer.

Table 1: LAN, MAN, and WAN

Internode distance	Node location	Example
0.1 meter	Circuit board	Data flow computers
1 meter	System	Multicomputer
10 meters	Room	LAN
100 meters	Building	LAN
1 kilometer	Campus	LAN
10 kilometers	City	MAN
100 kilometers	Country	WAN
1,000 kilometers	Continent	WAN
10,000 kilometers	Planet	Internet

The switching elements themselves do not consume the data bits, they just identifies the most efficient path and forward the data bits. These switching elements are usually referred to as *routers*. Table 1 illustrates the difference among LAN, MAN, and WAN with references to distance among computers connected by them.

Communication Topology

In these networks, two nodes exchange information using *point-to-point* (store-and-forward) or *packet-switched* scheme. In this scheme the entire information is received by an intermediate router, which forwards it to another router. This store-and-forward operation continues until the information reaches the destination node. There are a number of network topologies used to manage this kind of communication. These are (a) Star, (b) Ring, (c) Tree, (d) Complete, (e) Intersecting, and (f) Irregular.

Network software architecture

Information exchange between any two nodes goes through a number of steps before it is completed. To reduce the design complexity the entire process is divided into an *n-layer* architecture where each layer is built upon the one below it. Thus, in reality the data from a sender goes through these layers before it is sent to the receiving node. Similarly, at the destination node the data goes through these layers before it could be used.

The two main software architectures, which we refer to are (a) OSI (Open System Interconnection) reference architecture and (b) TCP/IP (Transmission Control Protocol/Internet Protocol) reference architecture.

Layered Structure

To increase modularity and availability and reduce complexity of inter-process communication the entire communication phase is divided into several layers. The entire structure is divided into two parts: *Logical* layers and *Physical* layers. In logical layers the processing demands are identified and the data to be transmitted is reformatted incrementally before it is sent to the destination node. The physical layer works on the formatted data (bit string) and transmits it to the other nodes.

The two-layer structure is enhanced (Figure 2) to show the position of OS. It illustrates the relative position of the functions of an operating system in the computer network.

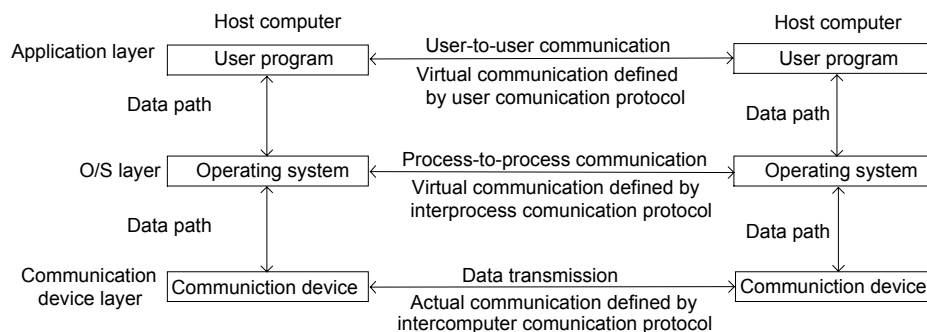


Figure 2: A three-layer structure

This three-layer structure is further elaborated by international standardization bodies and various vendors.

ISO/OSI

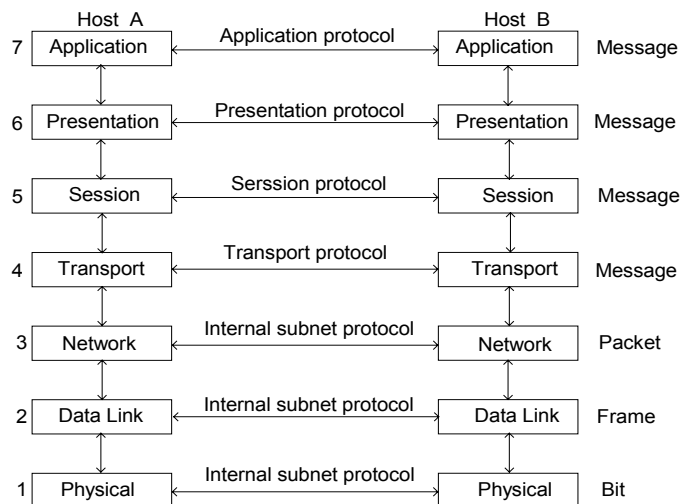


Figure 3: The Network architecture based on the ISO/OSI reference model

3. Different layers can be assigned to different standards committees, or different design teams.
4. Fundamentally different mechanisms may be substituted without affecting more than one layer (e.g., packet switching vs. leased line concentrators).
5. Different machines may plug in at different levels.
6. The relationship between the different control functions can be better understood when they are split into layers. This is especially true with the control actions, which occur sequentially in time from layer to layer.
7. Common lower level services may be shared by different higher level users.
8. Functions, especially at the lower layers, may be removed from software and built into hardware or micro code.
9. Plug compatible connections between machines of different manufacturers are made easier to accomplish.

Disadvantages of Layered Architectures

1. The total overhead is somewhat higher.
2. Two communication machines may have to use certain functions, which they could do without.
3. To make each layer usable by itself there is some small duplication of function between the layers.
4. As technology changes (e.g., as cryptography and compaction chips become available, or these functions can be built on to VLSI chips) the function may not be in the most cost-effective layer.

The Physical Layer (Layer 1): To transmit message (modified bit string) from sender host to the receiver host. This layer guarantees that if a sender sends bit 1 the receiver receives it as 1 and not 0.

Data Link Layer (Layer 2): A message in its raw form is not suitable for transmission. The data link layer does the following:

1. Transforms the raw bit string into dispatchable data frames by using special bit pattern as frame boundaries.
2. Transmits the frame sequentially.
3. Processing the acknowledgement frames sent back by the receiver.

An interference on the transmission line (noise) may destroy a packet during transmission. In this case, the layer 2 retransmits the packet. A retransmission of a packet may also occur if the acknowledgement is destroyed. It is the responsibility of layer 2 to resolve the problems caused by duplicate transmission, packet lost, error in transmission etc. Synchronization problem arises due to the speed difference in the sender and receiver. If the receiver is slow in receiving messages then the transmitter must control the transmission speed to match the receiver and vice versa. This is typically the producer consumer synchronization problem and layer makes sure that the synchronization is achieved correctly.

The Network Layer (Layer 3): The network layer is mainly responsible for routing and congestion control. A message can be routed either by the use of virtual circuit or by datagram approach. In a virtual circuit the network first establishes a communication route by sending a

setup packet, and once a route is setup the entire message communication (the subsequent packets). This approach is better suited for longer communication, for example, a file transfer that must be broken into a sequence of packets. In this way virtual circuit method is a static method. The datagram approach, every packet is treated individually, i.e., each packet contains the destination address. No prior route is established and each packet's route is found dynamically. In this way each packet may reach destination in a different route and also in a different order than the order in which they are sent. This method is well suited for short communication, i.e., electronic mail.

Layer 3 also maintains the accounting information for billing purposes. This may become complex when the message communication takes place between two countries.

The Transport Layer (Layer 4): This is also known as *host-to-host* layer, and serves as a boundary between the application-oriented layers (5 through 7) and communication-oriented layers (1 through 3). It permits many users to share the same transmission facilities in such a way as if each user has a simple point-to-point link. The main functions are summarized as follows:

1. Establishes a link to a remote machine. The logical link may employ multiple physical links.
2. Disconnects the logical link after use.
3. Delivers message over the logical link. Ensures that no message is lost. If a message or a part of message is lost or undeliverable, notifies the sender or returns the message to the sender.
4. Permits the receiving machine to regulate the rate at which it receives messages so as to avoid overloading.

The Session Layer (Layer 5): This layer provides the facility to support and maintain sessions (a connection between users, i.e., two processes is called a session) between two application entities. Sessions may span a long time interval involving many message interactions or be very short involving one or two messages. Typical sessions would be a terminal logged on into a remote computer and using the editor, or a transaction between a banking terminal and the bank's computer. In the case of loss of a communication connection, the session layer may provide checkpointing facility, to allow recovery to a known state. It may also provide facilities for controlling dialogue (half duplex, full duplex etc.).

In most distributed systems the session layer is minimal and is often incorporated within the transport or application layer.

The Presentation Layer (Layer 6): The purpose of the presentation layer is to resolve differences in the information representation between application entities. It allows communication between application entities running on different computers or implemented using different programming languages. This layer is concerned with data transformation, i.e., structuring, encryption, and compression. For example, it is necessary to provide a standard representation for integer, real and character codes for communication between heterogeneous computers. It would also be responsible for mapping more complex data structures such as list, trees and structure employing pointers for transfer between computers. Many of these functions are application dependent and are often performed by high-level language compilers, so the borderline between presentation and application layer is not clear.

The Application Layer (Layer 7): This layer is concerned with higher layer functions, which provide support to the application or system activities. Application level issues are:

1. Decomposition of functions into distributed processes: The application is a configuration of interconnected, communicating processes. Processes provide a modular software component from which to build the system. The criteria used for deciding where in the distributed system a particular process should run is usually for functional, geographical or performance reasons, i.e., functionally associated processes can be grouped together, or a process is allocated to a station closed to where it is used, or it is allocated to a station which is under utilized (load sharing),
2. Management of application processes: The overall management and the control of application can be organized in either a distributed or centralized manner.
3. Data distribution: The data in a distributed system can be organized in different ways.
4. Interprocess communication: The application processes communicate in order to cooperate and synchronize their activities.

Implementation strategies

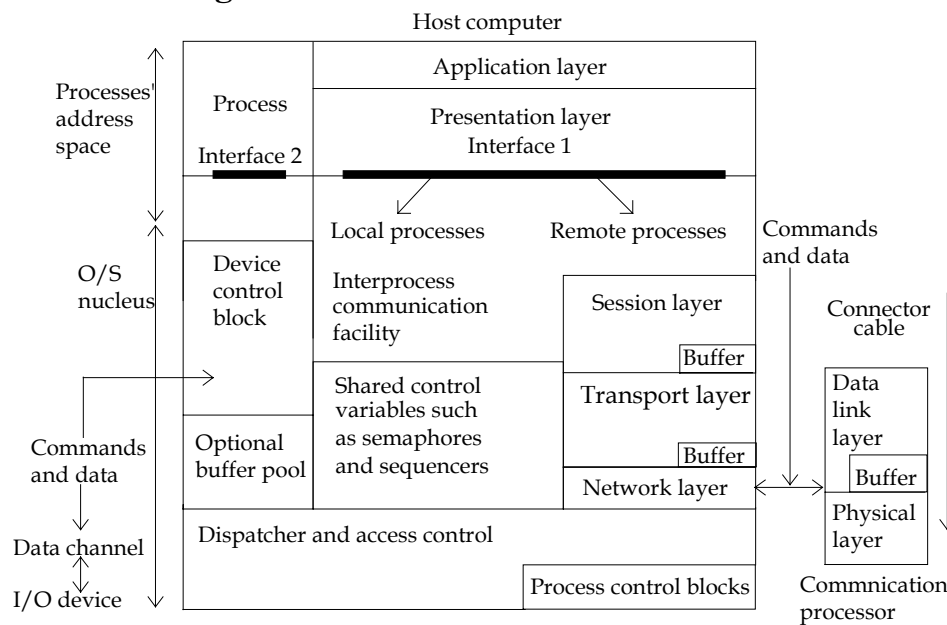


Figure 5: A possible implementation

The number of layers and division of duties among the layers vary from implementation to implementation. Figure 5 shows one possible implementation. A user process may communicate with either local or remote processes. Interface 1 directs a call to either the interprocess communication facility or the session layer, depending on whether the target process is local or remote. Interface 2 provides I/O services. The network layer of the host computer could be moved to a communication processor.

TCP/IP Architecture

TCP/IP is a four-layer communication architecture that can be conceptually described as a hierarchy built on a physical network interface that provides communication interfaces to the

network hardware. Figure 8 illustrates what ISO layers correspond to what TCP/IP layers. A communication protocol is an agreement that specifies a common language two computers used to exchange messages. A computer communication protocol defines communication precisely. It is a set of rules, which coordinates the exchange of messages between a client and a server. It specifies the exact format and meaning of each message that a computer can send. It also specifies the conditions under which a computer should send a given message, and how a computer should respond when a message arrives. In a radio communication between two parties, each party uses the key word "OVER" to end the message to be sent to the other party. The use of "OVER" in this communication session is a protocol. Data communication has extensive requirements, therefore, the process requires complex protocols. It involves a large number of steps. These steps could be combined into a single step, however, it is not reliable and inefficient to execute such a big step as a unit. For this reason small individual steps are handled by one protocol. A data communication originates at the application layer and finishes at the physical layer. Between the application layer and the physical layer, there are five more layers in the OSI reference model and there are only four layers in TCP/IP. Each layer is managed by a protocol.

TCP/IP is a collection of protocols, which are necessary for data communication. The main characteristics of TCP/IP are

- Connectionless protocols in the network layer.
- Nodes as packet-switching components.
- Dynamic routing transport protocols with security functions.
- A common set of application programs.

The Internet Protocol - IP: Internet Protocol (IP) is the corner-stone of TCP/IP architecture. It defines the basic unit of data transfer and the exact format of all data as it passes across the internet. It specifies how data packets should be processed and how errors should be handled. Each computer that connects to the internet must follow the rules of IP. When it creates a packet, a computer must use the format IP specifies. When a computer receives a packet, the packet will be an exact copy of the packet that was originally sent, still in the IP format. Furthermore, each router in the internet expects packets to adhere to the IP format as they pass from one network to another. Computer hardware does not understand IP. Therefore, attaching a computer to the internet does not mean it can use internet services. To communicate on the internet, a computer needs IP software. All internet services use IP to send or receive packets. Each computer usually has a single copy of IP software that all applications share. On sophisticated computers, the operating system keeps a copy of the IP software in memory all times, ready to send or receive packets. The basic unit of transfer is referred to as IP Datagram.