Chapter 7

Local Area Networks - Fundamentals

A Summary...

Basic concepts of Local Area Networks. Network topologies and characteristics. Resolving contentions for use of network media. The seven layer ISO/OSI model. Public Data Networks.

Read This Chapter If...

- You want to learn the fundamentals of computer networks
- You need to know about the function and application of Local Area Networks in the industrial environment.
- You want to know about the framework in which Local Area Networking standards are established.

7.1 Local Area Network Concepts

A data network is a mechanism by which many computer-based devices (referred to as network nodes) can communicate with one another on an "any node to any node" basis. A Local Area Network (LAN) is so named because the nodes on that network are located within a reasonable proximity (less than a kilometre) of one another. A point to point link between two nodes can therefore be considered as a network with two nodes and it shares many of the characteristics of larger networks.

We already know that even a simple, point to point serial communication link between two nodes needs to have many rules of protocol resolved before it can function correctly. We need to use the common signalling techniques, common character representations, complementary communications hardware and so on. We need the communications interchanges between nodes to be strictly governed by these rules of protocol so that conflicts can be resolved. All these issues are also true of networks - except that now we need to coordinate the communications between many nodes simultaneously.

The majority of networks use serial communication between nodes. With this in mind, there are a number of ways in which we can physically interconnect these nodes so that any one node can communicate with any other node in a serial form. Three of these interconnections are shown in Figure 7.1.

In Figure 7.1 (a), a switching unit (multiplexer) physically connects the serial cable from any one node to the serial cable from any other node. Once the switching unit has made a physical connection, then two nodes can communicate with one another as though the multiplexer was not present. Any node connected to the multiplexer can request it to make a physical connection with a target node. The multiplexer therefore needs to have some "intelligence", particularly if a conflict (contention) situation arises, where several nodes request connection to the same target node at the same time.

In Figure 7.1 (b) all the nodes are physically connected together on a central cable (trunk). Whenever one node places data on the trunk cable, all the other nodes receive that data. Since there is no intelligent switching device in this network, it is necessary to establish a system (protocol) where each node can recognise relevant data and ignore irrelevant data. This situation is the serial analogy to the parallel data bus structure within a computer system. Although it is possible to have multiple channels existing on a cable (through the use of modulation), a contention still arises if two devices attempt to transmit at the same time on the same channel. Each of the devices on such a network must therefore have the intelligence to independently react to and resolve such contention situations.

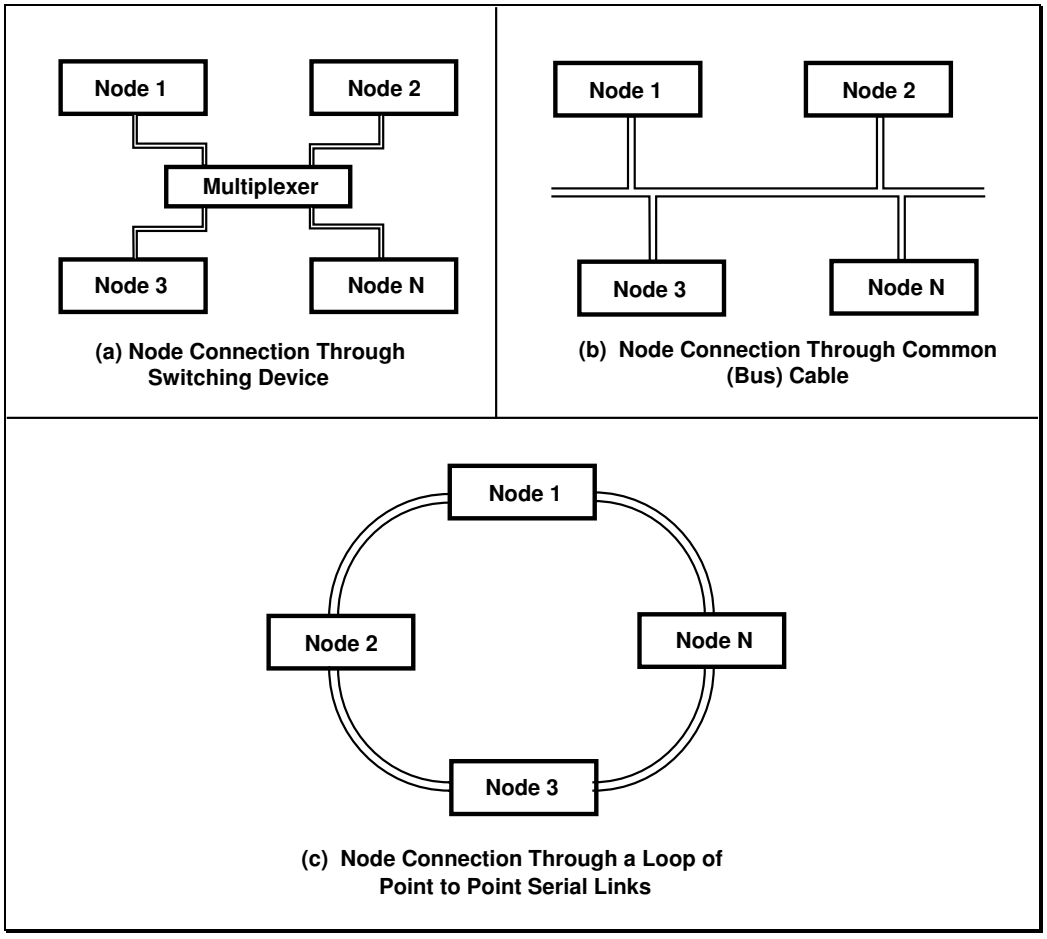


Figure 7.1 - Interconnection Techniques for Creating Data Networks

In Figure 7.1 (c), nodes are connected to one another via "point to point" serial links that together form a logical ring. Data is passed from one node to another around the ring. As in the system of 7.1 (b), there is no intelligent "coordinating node" and hence all nodes must have the intelligence to cope with contentions and recognise appropriately targeted data.

In a point to point serial link there is only one possible destination for all transmitted data - that is, the node at the other end of the serial link. However in a network, each transmitting node must theoretically be capable of sending information to any other node on that network. All the network forms therefore need to cope with the fundamental issue of "addressing". All nodes need to be given a unique "address" in order for data to be targeted correctly. The problem of network addressing is not unlike the problem of addressing devices (sharing the same data bus) within a computer system.

In the network of Figure 7.1 (a), a transmitting device needs to tell the multiplexer the address of the target node to which it wishes to be connected. The receiving device needs to know which device originated the message so that it can address a response. In the networks of Figure 7.1 (b) and (c) all devices have access to all messages. Therefore, a transmitter needs to know the address of the device to which it is sending a message. A receiving device needs to know its own address and the intended destination address of all incoming messages. Receiving devices in all systems must be programmed to only respond to messages that are specifically intended for them.

The need for addressing means that regardless of the physical network arrangement, data must be placed into suitable packets for transfer. Each packet of data moving through a network needs to contain some source and target addressing information. This enables a receiving device know which device is transmitting to it and where to send response messages. The concept of packet addressing is shown schematically in Figure 7.2.

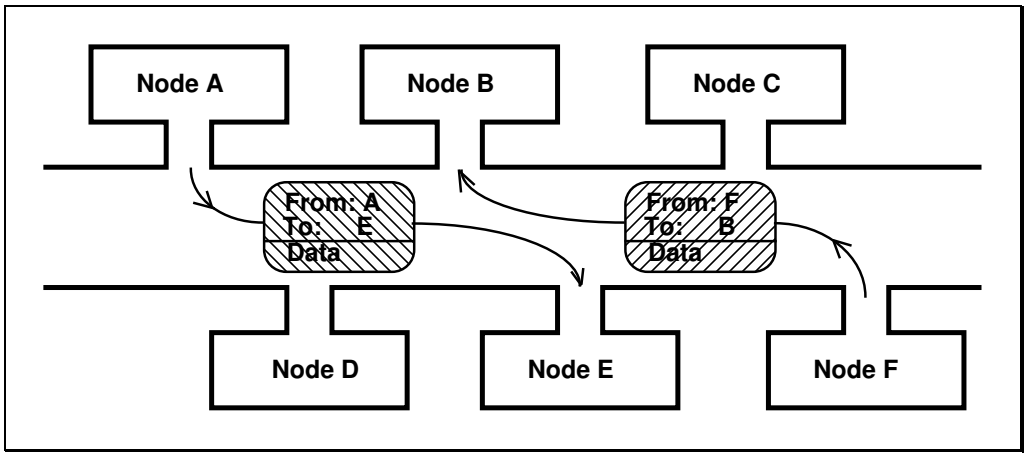


Figure 7.2 - Addressing Packets of Data on a Network

With the exception of traffic control (contention) and addressing functions, the issues related to networks are essentially the same as those in point to point links. We still need error checking mechanisms in the form of Block Check Sums (or more commonly, Cyclic Redundancy Checks). We still need hardware to perform parallel to serial (and vice-versa) conversion on each node. We still need layers of data handling software that can provide our applications programs with powerful communications sub-programs.

Another topic which sometimes causes confusion in regard to networks, is cabling. People often believe that there is something unique about the cables used in networks, as opposed to those used in point to point links. However, cable varieties used in common networks are the same as those used in point to point links, and include shielded-twisted-pair, co-axial and optic fibre. The major difference between network media and point to point media is in the connectors. If we choose to have a bus type network as in Figure 7.1 (b), then we need to have special connectors that can tap into the transmission medium at selected points. On the other hand, if we have a switched network such as that in Figure 7.1 (a), then we can often use the same connectors as in point to point communication - for example RS-232.

One of the most common LAN media, in the bus network arrangement, is the co-axial cable. One of the major reasons for this is because of the proliferation of low cost connectors and adaptors that resulted from the introduction of the American cable television system. Co-axial cable lends itself to "tapping" at any point so that many devices can be connected to a central trunk, as shown in Figure 7.1 (b). When co-axial cables are used in LANs, only one conductor is available for signal transmission (plus another common return line). Therefore if we wish to achieve full-duplex communications in LANs, we need to separate transmitted and received data into "forward" and "reverse" channels, through the use of modulation.

Twisted-pair cables are also extensively used in networking, but are more prone to electro-magnetic interference than co-axial cables. They are therefore better suited to the office, rather than the industrial environment. However, twisted-pair systems are low in cost and it is possible to simply have two pairs of wires for transmission and receipt of information - thus abrogating the need for modulation.

Optic fibre cables are currently not as prolific in LANs as they are in point to point links. The major reasons for this relate to the relative difficulty of tapping into optic fibres, and in terminating them without the use of specialised equipment. However, as optic fibre technology develops and the cost of connectors and adaptors decreases, it is evident that the system will become the dominant networking medium because of its superior noise immunity and higher bandwidth.

Cable types and connectors are not really a major issue in networking as far as end-users are concerned. In general, we have to accept a transmission medium from a limited range of commercial solutions that conform to an overall protocol. There are however, many other issues to be resolved when we attempt to physically interconnect a number of intelligent devices through a network. We shall look at these as we progress through the chapter.

7.2 Network Topologies

In section 7.1 we looked at three different ways in which nodes could be linked together to form a network. These physical arrangements are more commonly referred to as network "topologies".

Figure 7.3 shows each of the network topologies (introduced in the previous section) together with their common names.

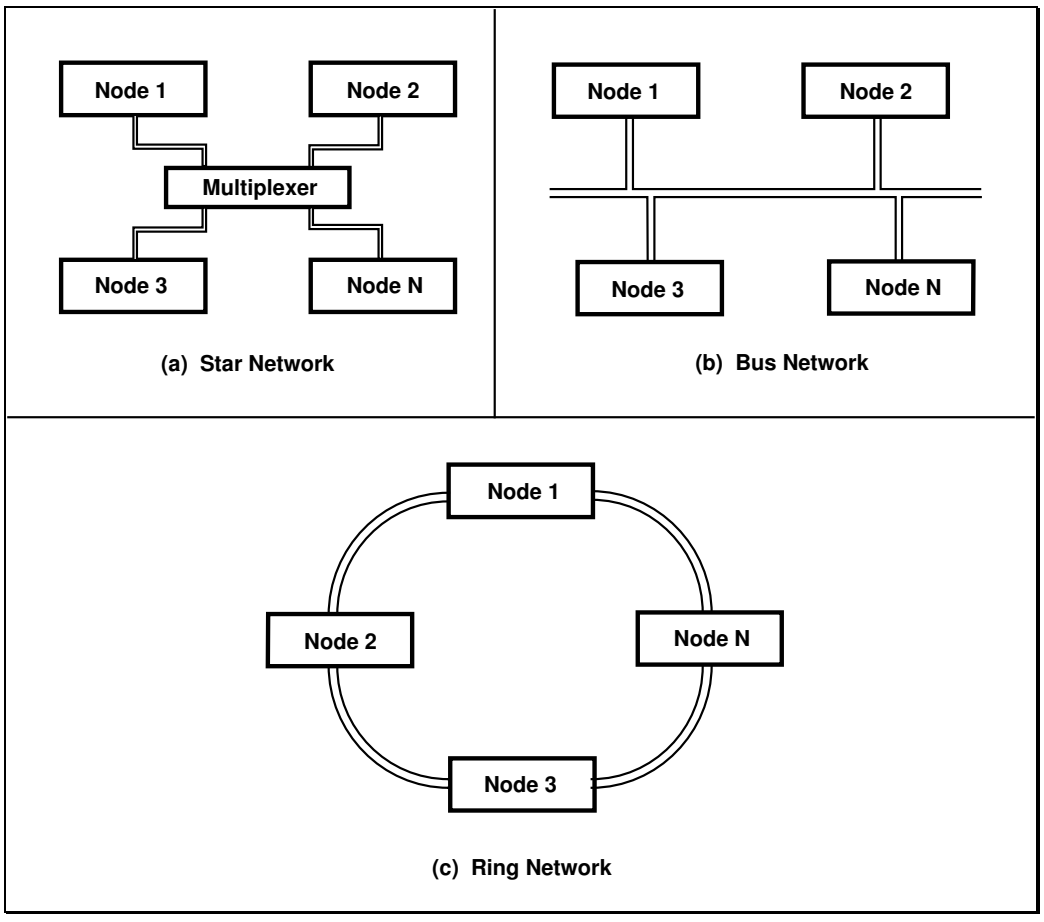


Figure 7.3 - Common Network Topologies

Each of the different network topologies has its advantages and disadvantages. There is no single solution that is optimal for all applications.

The star network has an intelligent central node, referred to as the "star node". The star node makes the decisions related to connecting any pair of nodes together. It therefore needs to be able to resolve any contentions that may arise. The star node is also responsible for tasks such as queuing requests for "link establishment" between nodes.

A star network is composed of a number of point to point links emanating from the star node. Several advantages stem from this. Firstly, the star node is transparent to communicating nodes once a connection has been made. Provided that the connected nodes agree on a software protocol, the nature of that protocol is of no consequence to the star node and the remainder of the network. In other words, devices 1 and 2 can talk to each other through "protocol A" and devices 3 and 4 can talk to each other through "protocol B". It is then also possible that device 3 can talk to device 1 through "protocol A". This has merits in manufacturing where it is not always practical to have all nodes using a single protocol and yet it may still be necessary to have all nodes capable of talking to one another. A good example of this would be where devices 1, 2 and 3 are computers and device 4 is a robot or CNC machine (with a fixed protocol).

Another advantage of the star network is that the physical medium used between any one node and the star node can be varied to suit the operating environment. For example, if device 1 is in the factory then it can be linked to the star node through an optic fibre cable. If device 2 is in the office, close to the star node, then it can be linked via a twisted-pair cable and so on.

There are also a number of disadvantages to the star network topology. All communication is dependent upon the star node - if it fails then all communications ceases. In a simple star network, the star node may be a microprocessor-controlled, serial port PABX (a multiplexer). In this case it is feasible to maintain some redundancy since the cost of the star node is minimal. At the other extreme however, a large star network could have a mini-computer as the star node, with many intelligent terminals communicating to one another through it. In this situation it is not practical to maintain redundancy.

Another shortcoming of the star network system is the cost of cabling. It is more cost effective to lay a single trunk cable through a networking area, and to use short tap cables from each node to the trunk, rather than to have long cables all meandering their way towards a central node. This is best visualised in terms of the power supply within a home. A power supply bus runs throughout a house and appliances tap into that bus at convenient points. It is a much neater solution than having every appliance connecting to a single, central point.

Another significant factor that arises in the cost equation for star networks is that of cable maintenance. In large star networks, many cables need to converge on the central node. Hence the concentration of cables, within ducts near the star node, is always very high. This makes trouble-shooting more difficult and time-consuming than it would otherwise be with other network configurations. In a large industrial environment, the problem of cable maintenance in star networks is severe. Often, defective old links are simply replaced with new links, and the old links left in the ducts, because the cost of removing them (or tracing problems in them) is so high. The "dead" links then further add to the cable concentration near the star node.

The problems of high cable concentrations are automatically eliminated in bus networks, where a central trunk of either twisted-pair, co-axial or optic fibre cable is laid down throughout a networking zone. Each node is connected into the network through a short tapping cable in a manner that is completely analogous to the domestic power supply scenario cited earlier.

The bus network is perhaps the most common form of the networking topologies - particularly in the industrial environment. A bus network is similar to the internal bus structure used for communications within a microprocessor system environment. The major differences are that in bus networks, data transfer is serial (not parallel) and secondly that there is no simple "master-slave" relationship between devices and therefore many contention situations can arise.

Bus networks offer a flexibility in terms of cable utilisation, which cannot be achieved with other network topologies. The fact that a bus network is based upon a trunk cable, which is laid throughout an entire area, means that video and voice channels can share the same cable, through the use of modulation techniques. This greatly increases the cost effectiveness of the bus network.

In a ring network, neighbouring nodes are interconnected with point to point serial links until a complete ring is formed. Data in ring networks is passed uni-directionally from node to node. Each device receives a message and then re-transmits it. This is shown in Figure 7.4.

A device in a network ring originates a message that is passed around the loop from node to node. Nodes in between the source and the destination do not alter the message. However, when the destination node receives the message, it modifies the control portion of the message packet and places it back onto the loop. The originator of the message packet determines whether or not the message has reached its target correctly by the modifications on the returning packet. Ring networks are relatively commonplace in the office environment, where the area they cover is relatively small. The response time of networks based upon the ring topology can be very good with an appropriate protocol.

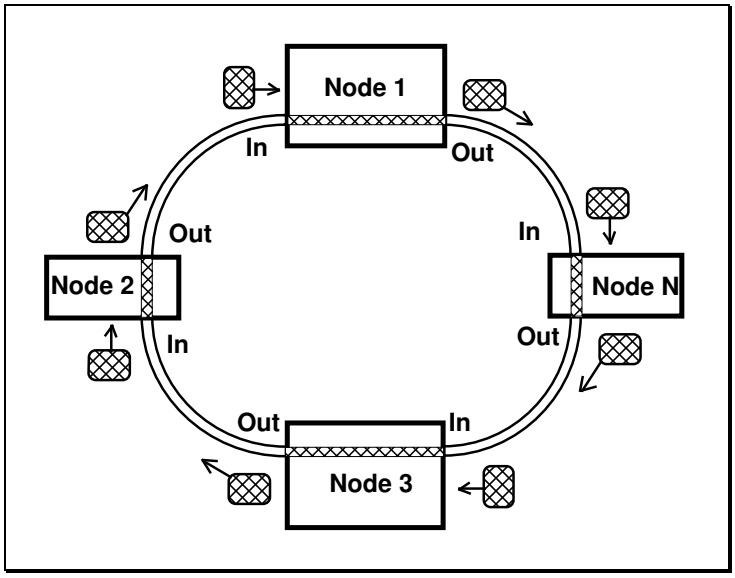


Figure 7.4 - Ring Network Message Transmission

A potential problem for the ring network topology arises because devices are all interlinked with point to point links. Hence one is tempted to ask what happens when a device fails - does the network stop? As it turns out, there are by-pass mechanisms built into ring networks so that devices that are down (or just switched off) provide a short-circuit path and do not result in network failure. However, the ring network completely fails if any one of the point to point links is severed.

In terms of cabling in ring networks it is evident that if one node is far removed from all other nodes then the cost of transmission medium will be higher than that in the bus network. For this reason, ring networks are most commonly found in the office environment for short-distance communications.

In theory, every node on both the bus and ring network topologies can have access to all the binary data that is transmitted through the communications medium. Nodes in these networks are normally selective about which data they respond to and base their response upon addressing information on data packets. However, from a more political point of view, bus and ring networks can be "tapped" into by unauthorised users. A tapping point anywhere within these networks will provide access to all network activity and information.

The bus and ring topologies are much more restrictive than the star in terms of the devices that can be connected into the network. All devices connecting into bus and ring networks must be capable of responding to a common communications protocol - otherwise such a network can clearly not function. In terms of general purpose (fully programmable) computer systems, it is feasible to link a wide range of devices to a bus or ring network. However, many devices simply do not fit into this type of communications structure. For example, few low-cost printers and many older industrial controllers cannot be interfaced into bus or ring networks, because they are not fully programmable and they do not have standard, internal bus structures.

Bus and ring networks offer one major advantage over the star network. To begin with, they are not dependent on a central, intelligent node to provide routing of messages and resolution of contentions that arise when a number of devices wish use the communications medium simultaneously. Each node in bus and ring network topologies is responsible for adherence to rules of protocol. In both bus and ring networks, the system can still continue to function even when one or more nodes become inoperative.

Ultimately the selection of a networking topology is governed by a number of factors including:

- the type of protocol selected to govern the network
- the availability of interfacing equipment
- the type of equipment being networked
- the environment in which the network is located.

It may be that, in the final analysis, the decision comes down to selecting the network to which the most devices can be interfaced with the minimum amount of effort.

In the manufacturing environment, the most common high-level network topology is the bus structure and the majority of protocols for this environment are based upon this topology. Star networks are also in widespread use in manufacturing to link CAD systems to a range of different CNC machines.

As with many engineering problems, there is no single solution that is correct for all applications. Sometimes the decisions that need to be reached are based on conformance or convenience rather than any technical reasons.

7.3 Contention Schemes

The technique of modulation allows us to utilise a transmission medium with a high degree of efficiency. Modulation is all about changing the physical representation of information, by creating communications channels, so that many different information systems can share the same transmission medium. This gives us the opportunity of using the same transmission medium for video, audio and digital data transmission - however, within any one channel conflicts can still arise.

In terms of data communications, each channel is generally restricted to one quantum of binary information at any instant in time. Sometimes this is for physical reasons and sometimes this is for practical reasons. For example, it may be impossible to decode a signal, which is the lump sum of a number of messages, back into the original (individual) messages. In a network, or more specifically in a bus network, we have a situation where there is the potential for many nodes to attempt to place data onto the transmission medium at the same time. This uncontrolled transmission could make decoding impossible. The physical conflict is called a contention situation.

Figure 7.5 shows a bus network in which devices are all tied together through a two wire conducting cable (signal + common line). Since all devices in this network are "intelligent", they are capable of placing data (represented by voltage levels) onto the bus at any time.

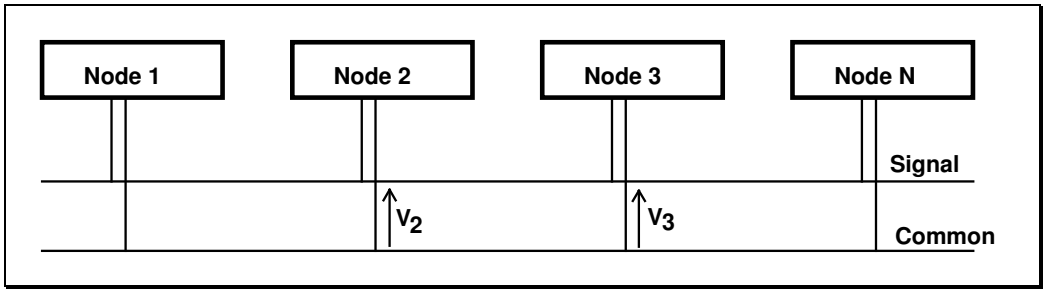


Figure 7.5 - Network with Devices Connected via a Conducting Bus

If we forget modulation for the time being (since we are considering communications within one channel anyway), we can consider the ramifications of two devices simultaneously trying to put binary data onto the line at the same time.

Consider the situation where both devices 2 and 3 put the same binary data onto the line - say a high voltage. There is no physical conflict, but nevertheless none of the other devices can detect that there are two devices speaking and not one. However, if device 2 places a low voltage on the line whilst device 3 places a high voltage on the line, then we have a physical conflict and a temporary short circuit on the cable between these devices. In practice this short circuit is probably not destructive, partly because the cable does have some resistance, but more importantly because the outputs and inputs of devices are buffered and do not have rigidly fixed voltage outputs. In any event the data that is now on the bus is meaningless.

Ultimately it doesn't really matter whether we are talking about transmission on a conducting cable or optic fibre cable or high frequency transmission through the air. The contention problem within any one channel always manifests itself in one energy form or another. It is important to note that the phenomenon of modulation does not cause the same physical conflict within the communications medium and that the individual elements within a modulated system can be recovered.

All of the above discussion is well and good, but what techniques are available to resolve the problem of contention on bus networks? Contentions can be resolved in any number of different ways, but there are two generic techniques of contention resolution that are in widespread use. These are the "CSMA/CD" system and the "Token Passing" system.

(i) CSMA/CD

CSMA/CD is an abbreviation for "Carrier Sense, Multiple Access with Collision Detection". The CSMA/CD system sounds complex but is straightforward to implement in practice. It is used within a number of different bus networks.

Each device in a CSMA/CD system is allowed to attempt to transmit on the network bus at any time. In other words, multiple access. However, prior to attempting a transmission, each device must monitor the bus for the presence of a carrier signal, emanating from another node. This is called "carrier sensing". If a carrier is already present on the bus (another node is already transmitting), then the device must wait until that transmission has ceased before attempting to place a message packet (frame) on the bus. Even when a device has the right to transmit, it must still monitor the bus to ensure that the signal that is being sent is the same as that on the bus.

It should be apparent, that two devices may simultaneously detect that the bus is clear and simultaneously attempt to transmit a message frame. In this case a collision will occur. Since all devices are monitoring the state of the bus, both of the devices will detect the collision. The first device to detect a collision must transmit a random bit pattern, referred to as a "jam sequence" for a short period of time. The jam sequence is designed to last long enough for the other colliding device/s to realise that a collision has occurred.

Since data is corrupted by a collision, both devices back-off for a random time interval and attempt complete re-transmission of message frames later. The CSMA/CD system is probabilistic in nature. In other words, there is no way of knowing how long it will take for a message to get from source to destination. A CSMA/CD network is therefore said to be non-deterministic.

The CSMA/CD system is not ideally suited to the industrial environment. The irony of CSMA/CD is that the time delay for messages is longest when the network is busiest and the network is generally busiest when abnormal or emergency conditions arise. Since we demand the fastest response time under emergency and abnormal conditions the network is often classified as unacceptable in the factory. CSMA/CD is however an excellent system for the office environment and is extremely efficient, where data transfer between nodes is sparse, and occasional time delays are of little consequence.

(ii) Token Passing Schemes

In principle, the so-called "token passing" scheme sounds much simpler than the CSMA/CD system. In practice it is more difficult to implement. The scheme is based upon a binary bit pattern that is referred to as a "token". Before any node is permitted to place message frames onto a network, it must be in possession of the token. Once a node has the token, it is permitted to transmit a message frame and must then pass the token on to another node. The movement of the token from node to node forms a "logical ring" between devices.

The token is itself a message frame (packet) with a special control section that defines its characteristics. A node wishing to use the token modifies these characteristics so that it can become a message frame. The node can then place data into the message frame. The token passing scheme is deterministic, because it is possible to precisely define the maximum delay that will arise in transmitting a data frame. It is for this reason that the scheme is often promoted as a basis for industrial networks.

There are a number of problems with the token passing scheme. Since it is possible for a device to fail while it is in possession of the token, steps must be taken to ensure that there is a mechanism for regenerating a lost token. This makes the system much more complex than the CSMA/CD system. The token passing scheme also introduces delays into the network even under light traffic conditions. In other words, a token is still passed from device to device, whether or not that device is to broadcast on the network. In the CSMA/CD scheme, a transmitting device can export information immediately if the bus is clear - with token passing the transmitter must always wait for the token.

Having introduced two of the more prolific contention resolution schemes, and the expressions "deterministic" and "non-deterministic", it is important to qualify their definitions. A network is the sum total of the communication medium, the contention resolution scheme and all the other software functions that go together to form applications support sub-programs. The fact that a contention resolution scheme is deterministic does not mean that the total network is deterministic. All the elements of a network need to operate in a complementary (deterministic) manner in order to provide a fixed response time.

7.4 ISO / OSI Seven Layer Model

In order to make a number of computer controlled devices communicate with one another, on a meaningful basis, there are many definitions that need to be established, including:

- Electrical signalling methods
- Synchronous/asynchronous transmission mode
- Bit rates
- Error detection and correction techniques
- Cable types and connector types
- Modulation techniques
- Contention schemes
- Software structure to support applications programs.

As it happens, there are an enormous number of available techniques, hardware and software systems that can be used to fulfil each requirement. It is little wonder then that there are so many different networks available.

Whilst variety may be the spice of life, it is a curse to those who endeavour to link computer based devices together. In the past, the major computer manufacturers tended work in different directions in regard to networking. Each manufacturer decided to choose their own specifications for each requirement in the networking process. Some computer manufacturers pioneered new techniques in particular aspects of networking - for example, in terms of software support modules or contention schemes. The end result was that during the 1970s and 1980s we experienced an explosion of different networks, many of which were proprietary in nature.

Since different computer manufacturers tended to specialise in different aspects of computing, users seldom had the luxury (nor the risks) of the single vendor environment. It also became apparent that no, single networking solution was universally applicable to all environments and hence a number of different standards would always be in existence. The obvious course of action was to find a way to rationalise the development of networking standards so that there may be some hope of creating links between different networks. Unless there is a systematic approach to network development, it is extremely difficult to link different networks together and to create low cost networking hardware.

The problem with trying to rationalise networking standards is that it is a very complex task to decide upon the scope of a standard for any individual networking requirement. That is, should a standard define just cables or perhaps cables and connectors or perhaps cables, connectors, signalling techniques and contention schemes. Where does one draw the boundaries for standards?

The International Standards Organisation (ISO) tackled this problem by developing a framework for what is referred to as "Open Systems Interconnection" or "OSI". The objective of this framework was to place all the requirements, for making a number of computers communicate with one another, into seven functional groups called "layers". The end result of this work was the OSI 7-layer Communications model that is shown in Figure 7.6.

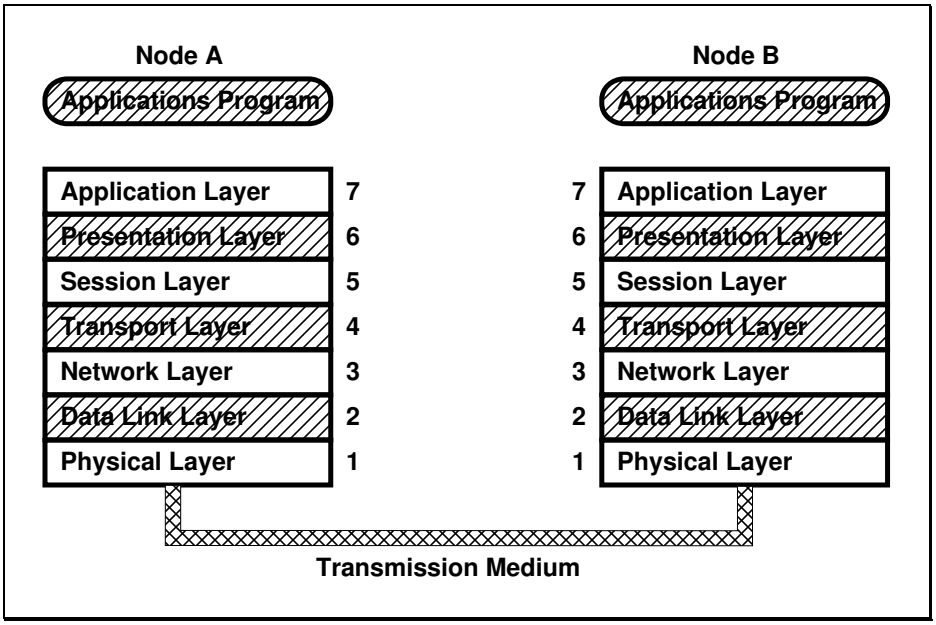


Figure 7.6 - Seven Layer ISO/OSI Model for Communications

It is important to note that the OSI model is not a networking standard in its own right. It is the framework, upon which, the development of communications standards is based. The choice of seven layers is somewhat arbitrary but is based upon many practical considerations. Each layer of the OSI model defines a number of related functions that must be enacted in order to take information from the layer below/above, process it and feed it to the layer above/below.

Some of the layers in the OSI model will naturally be more complex than others. However, it is necessary to understand that no, single layer is any more or less important than any other layer in realising a communication link. The layer representation is only used to show the logical sequence of steps for transmission or reception of information. The OSI model is therefore a seven layer shell in which the individual steps of the communications process are defined.

To illustrate the layering of communications tasks, let us suppose that computer A wishes to transmit a file to computer B through a network. The file must be broken down into a number of smaller units that can be sent as packets through the network. Each packet must be addressed to the target node. The source device, A, must contend for use of the network medium, access the network, transmit the information packets in either character or bit form and then transmit error checking information. If an error occurs then device B must request a re-transmission and device A must re-transmit appropriate information packets.

Without standardisation, the number of ways in which such a communications sequence could be programmed and implemented on each device is almost limitless.

If, on the other hand, one takes the layered approach to implementing a protocol, software and hardware can be developed in discrete modules. It is irrelevant how software or hardware is generated within these modules - provided that each module adheres to an accepted standard. The standard for each module (layer) defines the precise form in which data must be fed into a module, how it is processed and the exact form in which data is output from a module.

Schematically, the layer by layer packet (frame) assembly and disassembly that occurs during a transmission sequence (when an application program on computer A feeds data to the Application Layer for export to computer B) is illustrated in Figure 7.7.

In order to understand the concept of packet assembly and disassembly, it is sometimes helpful to examine a human analogy. Consider the situation where the director of large company A (Fred) wishes to send a simple note to the director of large company B (Harry). Fred sends a simple note:

Dear Harry
Please attend a meeting in my office at 10 a.m. tomorrow.
Regards, Fred

Fred passes the note to his secretary , who places the note into an envelope and addresses it. The secretary sends the note to the mail-room, where the people look at the address and decide upon the most appropriate mechanism by which the note could be sent - private courier (say). The courier decides upon the best medium for transferring the note (ie: air/sea/land). The note is ultimately transferred to the mail room of company B. The mail room extracts the address data from the envelope and decides that it should be sent to Harry's office. Harry's secretary receives the note, removes the envelope and places it into Harry's in-tray. Harry and Fred only see the raw data and don't concern themselves with any of the other issues involved in transferring the note.

Harry and Fred are analogous to the applications programs in the networking world. The two secretaries are analogous to the applications layers in the OSI model and so on. In company A, personnel are charged with the responsibility of building a package for transmission to company B. The personnel in company B are charged with the responsibility of disassembling the incoming package so that the original data can be recovered. Figure 7.7 therefore shows how the OSI model provides a framework for an analogous functionality in the networking environment.

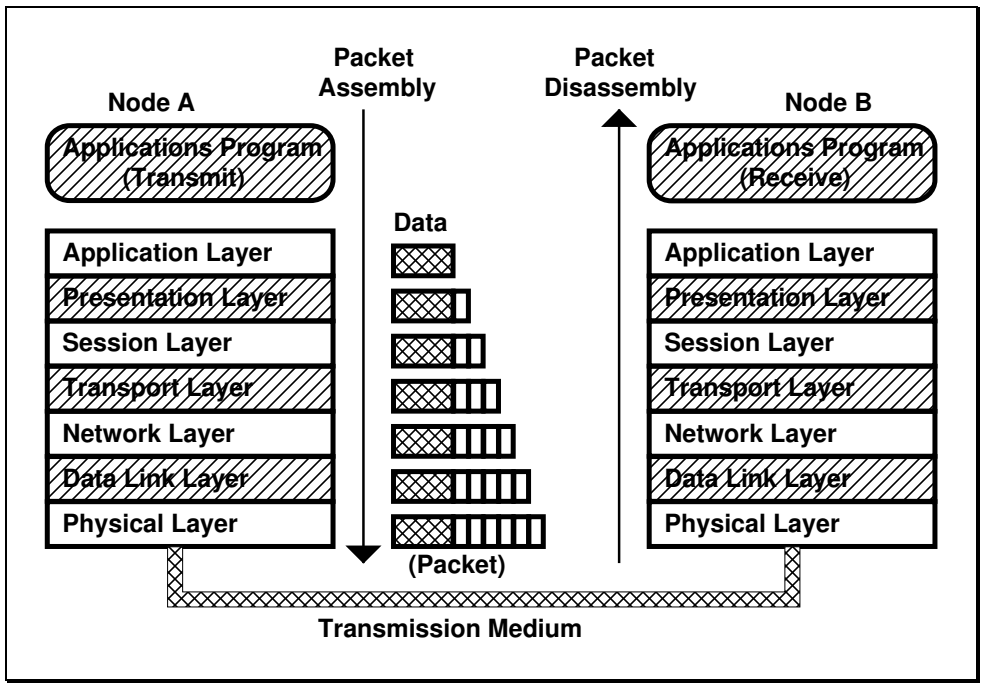


Figure 7.7 - Packet Assembly & Disassembly in OSI Networks

Provided that all the OSI layers have been put into place on a computer, then a user who wishes to program a device to transmit information through a network need only be concerned with the Application Layer. This layer contains all the sub-programs, related to network activity, which can be called up in any user's end application program. The actual role of each layer is briefly outlined below:

(i) Physical Layer

This layer is concerned with the physical connection between devices. It includes factors such as network topology, cable types, connector types, signal modulation types and contention schemes. It is the physical layer that is directly responsible for transmitting a stream of binary digits from one device to another.

(ii) Data Link Layer

The Data Link layer is responsible for ensuring the integrity of the bit streams that are transferred to/from the physical layer from/to the network layer. It is the data link layer that takes care of error detection and correction through the re-transmission of messages. Data link layers can provide both "connectionless" and "connection oriented" services. In a connectionless system, each information packet is treated as a self contained entity that is transferred to a target node without a two-way dialogue (connection) having been established. In a connection oriented system, devices try to establish a physical link before attempting data transmission. Whilst the physical layer puts the contention scheme into place, it is the Data Link Layer that is responsible for accessing the communications medium (Media Access Control).

(iii) Network Layer

The network layer is primarily concerned with message routing (or addressing) functions. It is designed to establish and clear logical or physical connections across the particular network in use. The network layer, to some extent, also has responsibility for flow control between devices. The network, data link and physical layers are all interdependent and hence the standards for each of these need to be selected with a view to forming a cohesive system based on all three layers.

(iv) Transport Layer

The transport layer is the one that interfaces the network dependent layers below to the network-independent, applications layers above. It is responsible for establishing a reliable message interchange service (between nodes) and providing this service to the session layer above. Since the transport layer's performance is restricted by the types of service available from the lower, three layers it must also be designed to provide different levels of service - referred to as "classes of service" or "quality of service". There are five classes of service, the highest being "Level 4" and providing complete flow and error control procedures. The lowest class of service is called "Level 0" and this only provides the basic functions required for connection establishment and data transfer.

(v) Session Layer

The session layer, as its name implies, is responsible for coordinating a communications session between nodes on a network. In other words, it establishes a logical connection between two nodes and controls the entire message interchange process that takes place between them during a communications session.

(vi) Presentation Layer

The presentation layer is the one that takes incoming data, which arrives in a common pseudo form and converts it to the form required by the application layer in the local device. Similarly, the presentation layer takes data from its local application layer and converts it into a common pseudo form for transmission. The pseudo form for data is referred to as the "transfer" or "concrete" syntax. The form in which data is presented and used within the application layer is referred to as the "abstract data syntax". Two communicating nodes may have different abstract data syntaxes.

(vii) Application Layer

The application layer is the one that provides direct support for applications software. The software developer has access to a set of "primitives" that transparently provide all the network services by interaction with the lower layers. The applications layer allows a user to call up these primitives and access information and files from remote nodes as though they were hardware within the local device. For example, the capture of a file from a remote device, on the network, is analogous to fetching a file from a local hard disk.

7.5 A Summary of Key Points Related to Networks

At this time it is appropriate to provide a summarised list of fundamental points related to Local Area Networks:

(i) A Local Area Network is a collection of intelligent devices (nodes) that are all interconnected via a star, bus or ring topology. In theory, any node should be able to communicate with any other node.

(ii) The majority of commercially available networks are based upon the use of serial communications techniques. Each node must have a conversion circuit to transform data from the parallel form, used on the internal data bus, to the serial form used on the network.

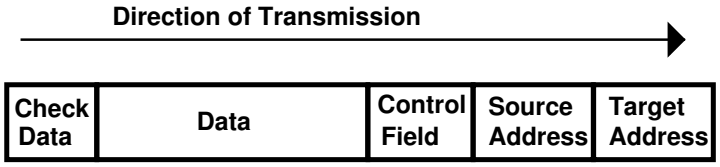
(iii) Many commercially available networks operate at very high bit rates (5 - 10 Mbits/sec) and therefore use synchronous serial communications, because of the difficulties involved in using asynchronous communications at these speeds.

(iv) In some networks all nodes have access to all information. It is therefore important that each node be given an address, that each node is aware of its address and that all messages contain a source and target address.

(v) Data transfer in networks can either be bit-oriented or character-oriented, depending upon the communications protocol in use.

(vi) Networks generally use cyclic redundancy check polynomials for error detection, but may also use Block Check Sums where data transfer rates are lower and noise immunity is high.

(vii) All data that is transferred on a network must be in a packet (frame) form so that addressing and error detection information can be included with the data. The fundamental features of a data packet are shown below:



(viii) Within any one communication channel on the network, whenever two or more nodes attempt to use the transmission medium simultaneously, a contention occurs. Contentions can lead to data corruption and therefore a number of contention schemes have been put into place. The form of these schemes depends upon the nature and environment of the network, but the two most common are the CSMA/CD and the Token Passing schemes.

(ix) The three basic network topologies are the star, bus and ring.

(x) The star network topology is composed of a number of point to point links between peripheral nodes and a central, star node. The intelligent "star node" is used to connect any one peripheral node to any other peripheral node. The star node also resolves contentions for resources.

(xi) The Bus Network consists of a central trunk cable, which is generally a two conductor (signal + return) or two optic fibre media. Nodes on the network all have access to the same data and must selectively ignore or act upon data packets, depending upon the addressing. Since bus networks do not have an intelligent node to resolve contentions, each node must be capable of handling contention situations.

(xii) The ring network topology is made up of a number of devices that are connected via "point to point links" to form a physical ring. Messages are generally sent around a ring from a source node to a destination node. Messages usually only travel in one direction around the ring. Nodes on the network all have access to the same data and must selectively ignore or act upon data packets, depending upon the addressing. Ring networks do not have an intelligent node to resolve contentions.

(xiii) The International Standards Organisation (ISO) has defined a 7 layer model for Open Systems Interconnection (OSI) in data communications. The purpose of the model is to break the communications process up into discrete modules so that standards can be clearly defined for each module. The OSI model is to be considered as a framework for communications standards.

7.6 Data Packet Forms on Networks - BSC, HDLC and SDLC

We have already examined many aspects of the OSI model's physical layer. The issues that are covered by this layer include network topology, communications media, modulation techniques, contention schemes and so on. We shall now examine some of the data link layer (OSI layer 2) protocols that are in common use with various types of networks.

You will recall that the data link layer of the OSI model is concerned with maintaining the integrity of data that is transmitted between nodes on a communications network. There are a number of protocols that are commonly used to implement data link control. It is important to examine these in some detail, since the acronyms used in relation to the data link layer protocols are prevalent in networking literature. Some of these protocols are character-oriented and some are bit-oriented.

(i) Basic Mode (BSC) protocol

The Basic Mode protocol, defined by the ISO, is primarily designed as a character-oriented protocol for synchronous serial transmission. It can also be used to transfer bit-oriented data in what is referred to as "transparent mode". Basic Mode is more commonly referred to by its IBM given names "BiSync" or "Binary Synchronous Control" or "BSC".

Under the Basic Mode protocol, each block (packet) of information is preceded by sending two, or more, synchronising characters, known as "SYN". After a prolonged idle period, a receiver synchronises itself to a transmitter when it detects the bit patterns of these synchronising characters. Since the BiSync protocol is commonly used with IBM computers, characters are usually represented in EBCDIC form. However the protocol also allows for ASCII and Six Bit Transcode representation.

The BiSync protocol allows for single-block or multiple-block messages, with or without headers (containing addressing information). This is because the protocol can be used for both point to point and multidrop (network) applications. A typical single-block message is shown in Figure 7.7. However, in a multiple-block message, the data in Figure 7.7 would be terminated with an "ETB" (End of Transmission Block) character, rather than an ETX character, to signify that more packets are to follow.

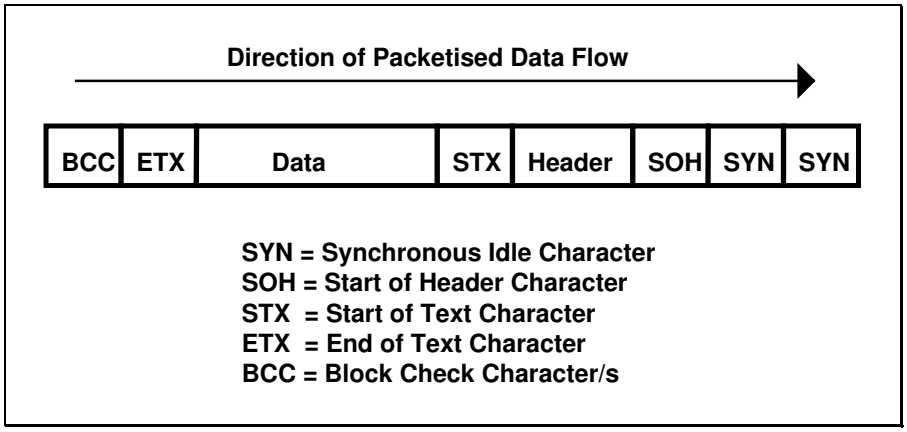


Figure 7.7 - Typical Block (Packet) in BiSync Data Link Protocol

The BiSync protocol also defines the following packets:

- ENQ (Enquiry)
- ACK (Acknowledge)
- NAK (Negative Acknowledge)
- EOT (End of Transmission)

These consist only of the single communications characters plus the synchronising characters (SYN).

The BiSync protocol is then essentially a synchronous ACK/NAK type protocol, with a receiver performing the same block check calculation on data as the transmitter. When a locally calculated block check character is the same as the block check character received from a remote device, a receiver returns an acknowledge packet - otherwise a negative acknowledge packet is returned to the originator of the packet.

If the data transmitted in a BiSync block is bit-oriented rather than character-oriented, then it is possible that some of the control characters (DLE, STX, etc.) could coincidentally be contained within the data. The rule used in BiSync (and many other protocols) is that if a "DLE" bit stream coincidentally occurs in data then an additional DLE is inserted. A receiving device performs the same check on the data field of an incoming block and strips out additional DLEs where necessary.

The form of the check character/s in the BiSync protocol depends upon the form of the data that is being transmitted. When BiSync is used to transfer character-oriented data then a simple Block Check Character is used for error detection and correction. If however, the protocol is used to transfer pure binary data in, what is referred to as "transparent mode", then a 16 bit Cyclic Redundancy Check is used in place of the BCC.

The primary advantage of BiSync is that it is relatively simple to implement in terms of hardware and software. As a result of this, and its nexus with IBM, it has therefore gained a broad acceptance in the communications world.

One of the shortcomings of the BiSync protocol is that it is half-duplex in its operation. Since many network installations now provide cables with sufficient bandwidth for full-duplex transmission, BiSync effectively "wastes" the bandwidth that is provided for the return channel.

(ii) HDLC / SDLC

The International Standards Organisation has defined the "High Level Data Link Control (HDLC)", which is a bit-oriented, synchronous protocol. It is almost (but not exactly) identical to IBM's Synchronous Data Link Control or SDLC. Like BiSync, HDLC is a protocol for the data link layer of the ISO model.

The HDLC protocol allows for full-duplex communications on either a simple point to point link or a multidrop network arrangement. Under the HDLC protocol, data can only be transmitted within a packet defined by a standard format. In HDLC parlance, a packet is more commonly referred to as a "frame". The structure of the bit-oriented HDLC frame is shown in Figure 7.8.

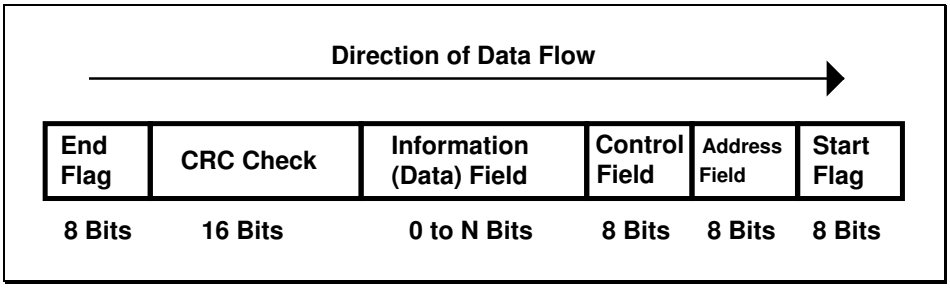


Figure 7.8 - HDLC Frame Format

The start and end flags for a HDLC frame are composed of unique bit patterns that are used to synchronise the receiving device. The integrity of each frame is checked through a Frame Check Sequence (FCS) which is a Cyclic Redundancy Check polynomial. The generator polynomial for the HDLC protocol is given by:

$$x^{16} + x^{12} + x^5 + x^0$$

An HDLC frame can be one of three types (classes):

- Unnumbered
- Information
- Supervisory

The frame class is defined by the bit pattern contained within the control field of the HDLC frame. Unnumbered HDLC frames are used for functions such as link set up and disconnect (analogous to Enquiry and Acknowledge in character-oriented protocols). Information frames, as the name implies, are the ones that carry actual data during an information transfer session. Supervisory frames are used for flow control and error checking purposes.

There are two modes of operation for an HDLC link. These are:

- Unbalanced Normal Response Mode (NRM)
- Asynchronous Balanced Mode (ABM)

The mode in which the link is to operate is defined during the initial set up of the link.

In NRM links, there is generally only one master (primary) station and a number of slave (secondary) stations. Secondary stations may only transmit when instructed to do so by the master station. This arrangement could correspond to a "mainframe to many terminals" arrangement. In ABM links, all stations have an equal ranking and therefore act in both primary and secondary roles. The ABM mode is most commonly used in point to point links.

A key feature of the HDLC protocol is the addressing of frames for a networked environment. Addressing is predominantly used for the NRM mode of link operation. Each secondary node on an NRM network can be given a unique address. When a primary node transmits a frame to a secondary node, then the address of the target node is placed into the address field of the frame.

It is also possible to provide a common address for a number of secondary nodes, so that all the nodes within the common address group receive a message from a primary node. This is referred to as "group addressing" and individual groups of secondary nodes can be specified. When the address field of an HDLC frame contains all ones, then the link is said to be in a "broadcast mode" and all secondary devices receive the broadcast frame from the primary node. An address field with all ones is referred to as a "broadcast address".

In principle, the HDLC link protocol is not unlike any other protocol and contains the normal transaction phases including:

- link establishment
- packet transmission
- error check transmission
- acknowledgment or negative acknowledgment of packets
- termination of transmission.

The HDLC protocol is however, relatively sophisticated because the control field not only contains link establishment and supervisory functions, but also frame sequencing information. The HDLC sequencing system allows for a number of frames to be transmitted to a target node without waiting for the target node to acknowledge each individual frame. A receiving device can therefore track the order of incoming frames and correct for those frames that arrive out of sequence.

7.7 PSTN / PSDN / CSDN / ISDN

Up until now we have only examined networking within a local area, be it in a single building or single factory. However, it is important from a communications point of view to be aware of the networks that link computer-based devices across the world. These Wide Area Networks (WANs) clearly have important consequences for large manufacturing and financial organisations where international data transfers need to occur on a regular basis.

Initially data communication between computer systems, separated by large distances, was carried out through the normal lines on the Public Switched Telephone Network (PSTN). Computers transmitted data to one another on these public telephone lines via modems. Unfortunately because of the channel bandwidths on public lines and the switching delays, data transfer rates were generally low (normally less than 4800 bps) and the timed charges imposed by the telephone companies were high. Many large computer organisations and financial institutions therefore chose to introduce their own private networks by leasing dedicated public lines from telephone companies and providing their own exchanges (switching nodes). These allowed companies to decrease the switching delays, maximise transmission bandwidth and therefore transmission speed.

Private data networks functioned well within a single organisation or within a single network, but problems arose when one organisation wanted to transfer its computer data to another organisation that was on a different, private network. It became evident that a public, wide area data transfer system, analogous to the switched telephone network, had to be introduced. This system is referred to as a Public Data Network or PDN. Public Data Networks are generally established and operated by a national administrative body in order to maintain standards. The traditional, switched telephone network is still widely used for data communications, and therefore, it too is a PDN.

There are essentially two different forms of Public Data Networks. These are the "Packet Switched Data Networks" (PSDNs) and the "Circuit Switched Data Networks" (CSDNs). The standards that are used in conjunction with both these types of Public Data Networks are those which occupy the "Network Layer" of the OSI 7-Layer model. A circuit switched network is one in which a group of intermediate exchanges set up a direct physical (electric circuit) connection between a transmitting device and a receiving device by short-circuiting appropriate incoming lines to outgoing lines. The circuit remains connected for the duration of a transaction (call). In other words, in an ideal environment, the transmitter and receiver are linked by a lossless cable. This is shown schematically in Figure 7.9.

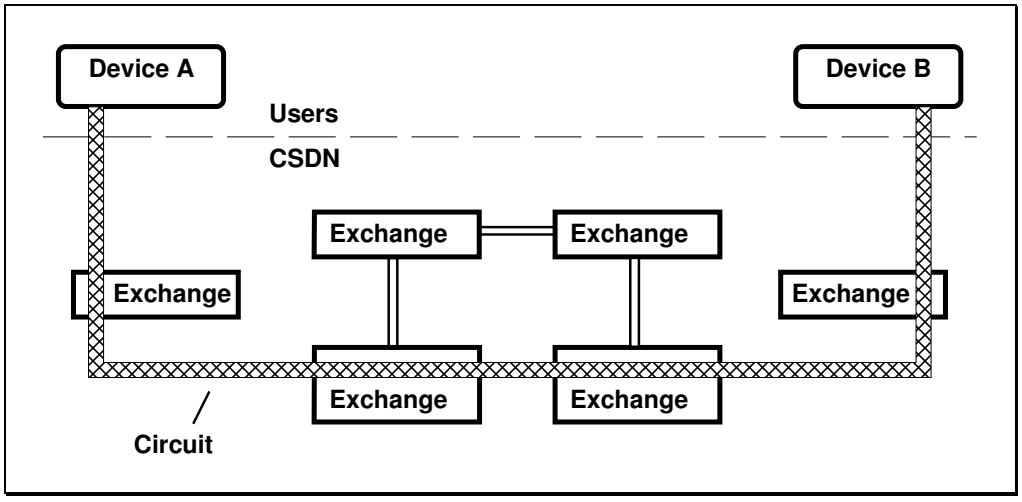


Figure 7.9 - Circuit Switched Data Network (CSDN) Operation

The PSTN is a good example of a circuit switched network, where exchanges perform either mechanical or electrical switching of transmission lines.

In a packet switched data network, all messages (regardless of length) are divided into discrete units (packets), which are transmitted from a source station to a destination through intermediate exchanges. Each packet contains a source and target address that intermediate exchanges use for routing the packet. A transmitter sends a packet to its local exchange. The local exchange reads the destination address and uses its "routing directory" to determine the next exchange to which the packet must be sent. Each exchange is said to perform a "packet store and forward" operation. Unlike the CSDN there is no direct, physical connection or communication between a transmitter and receiver - only between consecutive exchanges. This system is shown schematically in Figure 7.10.

Many transmitters can access the same exchange and hence the exchange will not necessarily forward packets in the consecutive order in which they are received from any one transmitter. The data travelling from one exchange to another is generally a mixture of packets from different transmitters. The maximum length of any one packet is restricted by the network protocol and hence no single transmitter can block the network with long messages.

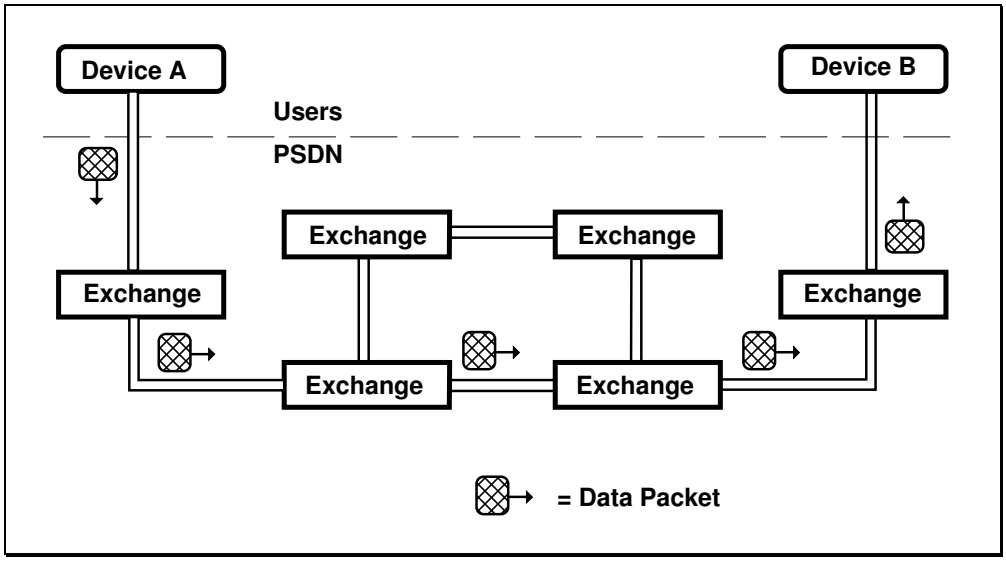


Figure 7.10 - Packet Switched Data Network Operation

When packets are transmitted as individual entities, as described above, the packet switched network is said to be operating in a "datagram" mode. However, it is also possible for the exchanges to set up packet transfer so that two communicating devices believe they are talking to one another through a physical circuit. This is referred to as a "virtual call" or "virtual circuit" mode of operation. Although the two user devices are never directly communicating, the exchanges make it appear as though this was the case.

The widely adopted standard for the format of a packet on a packet switched network is the CCITT X.25 standard. You will come across this standard regularly when reading material related to Packet Switched Wide Area Networks. X.25 defines the rules for connecting terminals to a packet switched exchange system, the format of the network packet header (containing addressing information) and the data section of the packet. The data-link layer of the OSI model treats the total X.25 packet as the data segment for its frame format. For example, the X.25 packet can form the data segment for the HDLC data link frame. This is shown in Figure 7.11.

The very nature of a packet switched data network indicates that error detection and correction on each packet must be performed after every transfer segment from one exchange to another. This error checking and correction is performed over and above any error checking carried out by the data link layer. In the circuit switched system, the exchanges act only as switches and so error detection and correction are performed by the two communicating users only, through the data link layer.

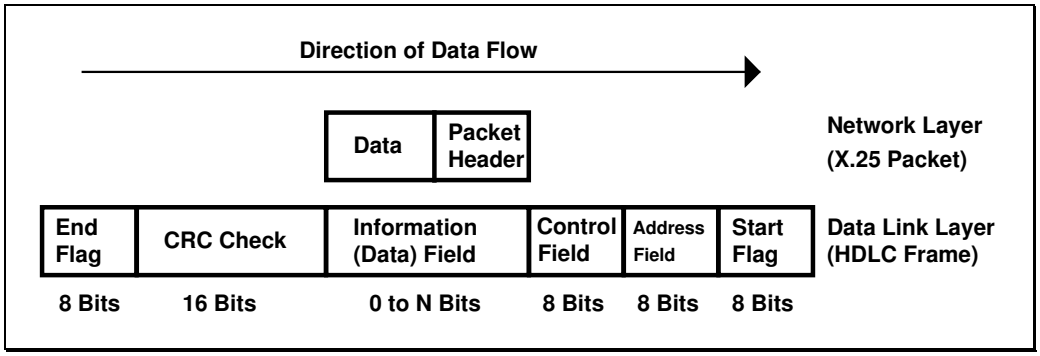


Figure 7.11 - Interaction of Network and Data Link Layers for Packet Switched Networks

While the data-oriented CSDN and PSDN serve the needs of computer users extremely well, the Public Switched Telephone Network is also regaining acceptance as a viable form of computer network. This has been largely due to the replacement of mechanical exchanges with digital exchanges that dramatically decrease call set-up times and provide a digital interface to all users. Once a total digital exchange system becomes fully operational on an international basis, there are many benefits to be attained, in a system that combines voice and data communications. The combined system is referred to as an Integrated Services Digital Network or ISDN. The digital interface to ISDN allows users to transfer data over the links, without the use of modems, at moderate bit rates (64 kbps) which were not available with traditional PSTNs.

7.8 The Role of Networking in Manufacturing

We have now examined some of the basic concepts of computer networking, but how do these complex structures fit into our perceptions of the manufacturing industry? Despite the fact that computer-based systems have been an integral part of manufacturing since the 1960s, networks sit very uncomfortably with manufacturing organisations. The reason for this is that the focus of engineering within manufacturing has changed dramatically with the introduction of specialised computer-controlled production equipment. The availability of low-cost microprocessors has led to a proliferation of intelligent, electronic control and feedback systems for mechanical devices. Until recently, there have been few professionals who have been able to successfully cope with the heavy demands of unifying these two, complex engineering disciplines.

The advent of intelligent manufacturing equipment has however caused professionals in industry to re-evaluate traditional manufacturing methods and consider the concept of Computer Integrated Manufacture (CIM). CIM is little more than the rationalisation and automation of the collection and distribution of computer-based manufacturing data. One of the major benefits of CIM is that the efficiency of an organisation can be readily monitored and its response time to changing demands minimised. In terms of computer networking, CIM can be visualised through the idealistic, "plug-in-compatible" network shown in Figure 7.12.

It would appear that the introduction of computer controlled equipment is therefore an ideal catalyst for CIM, but in the real world, CIM is far easier said than realised. On the other hand, if it was possible to simply plug each piece of computer controlled manufacturing equipment into a standard Local Area Network ("plug-in compatibility"), then there would at least be a sound basis for integrating and utilising manufacturing data so that we could ultimately optimise plant efficiency. However, we now know that the concept of "plug-in compatibility" is difficult to achieve because of the enormous diversity of computer control architectures and equally diverse needs within the manufacturing environment.

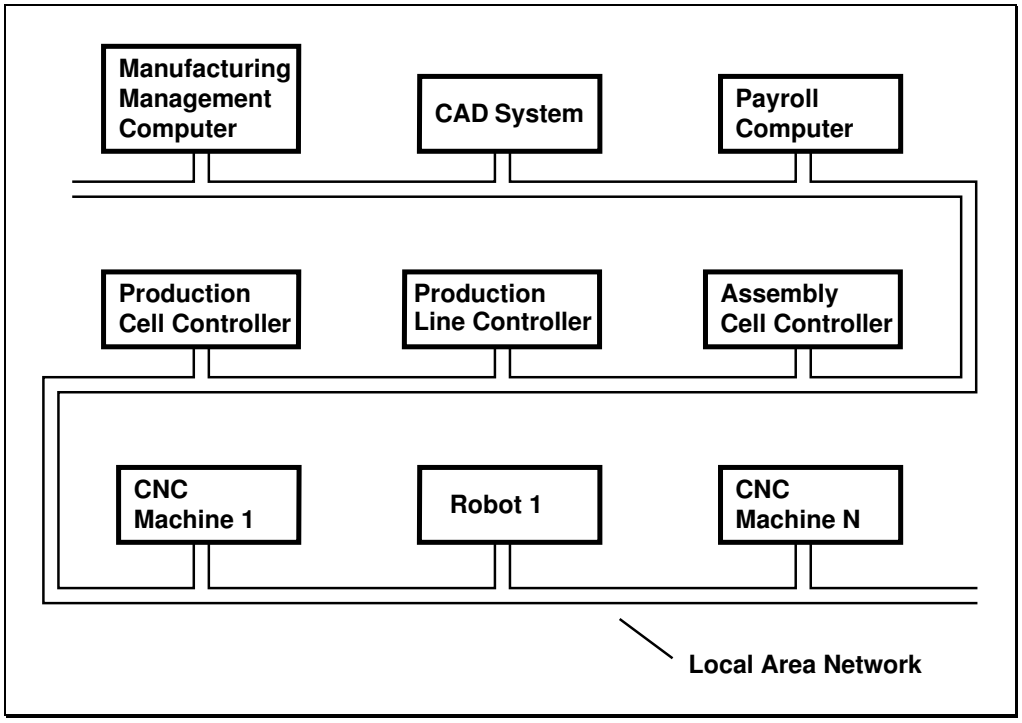


Figure 7.12 - Achieving CIM Through "Plug-in Compatible" Networks

In this chapter we have examined just a few of the many standards that can be used to fill the framework of the OSI model. As we have seen, there is no single, optimal solution to the networking problem. The network performance requirements of the office environment and the factory environment, even within a single manufacturing organisation, can be so different, that it is often impractical to implement the single network solution of Figure 7.12.

During the early 1980s, the problems of interfacing a diverse range of industrial controllers to a single network were so large that the original "nebulous" concept of CIM became little more than a pipe-dream. The ugly reality of CIM often began to resemble the ad-hoc sort of networking structure of Figure 7.13.

1980s CIM became a jumbled mixture of commercial, proprietary networks in the office environment, proprietary RS-232 ACK/NAK protocols between CAD and CAM and proprietary bus type networks between industrial control systems. Whilst the ideal network of Figure 7.12 was not feasible in the 1980s, it was also apparent that the proprietary systems were totally unacceptable for the long term.

At best, the proprietary networks in the factory provided interfacing units for only a limited range of industrial controllers (made by the larger manufacturers of automation control equipment). Smaller manufacturers of automation controls just didn't fit into the CIM picture and nor could they afford to, since there was no single, standard for industrial networking. The cost of an original equipment producer manufacturing interfaces to the entire range of differing proprietary networks is prohibitive.

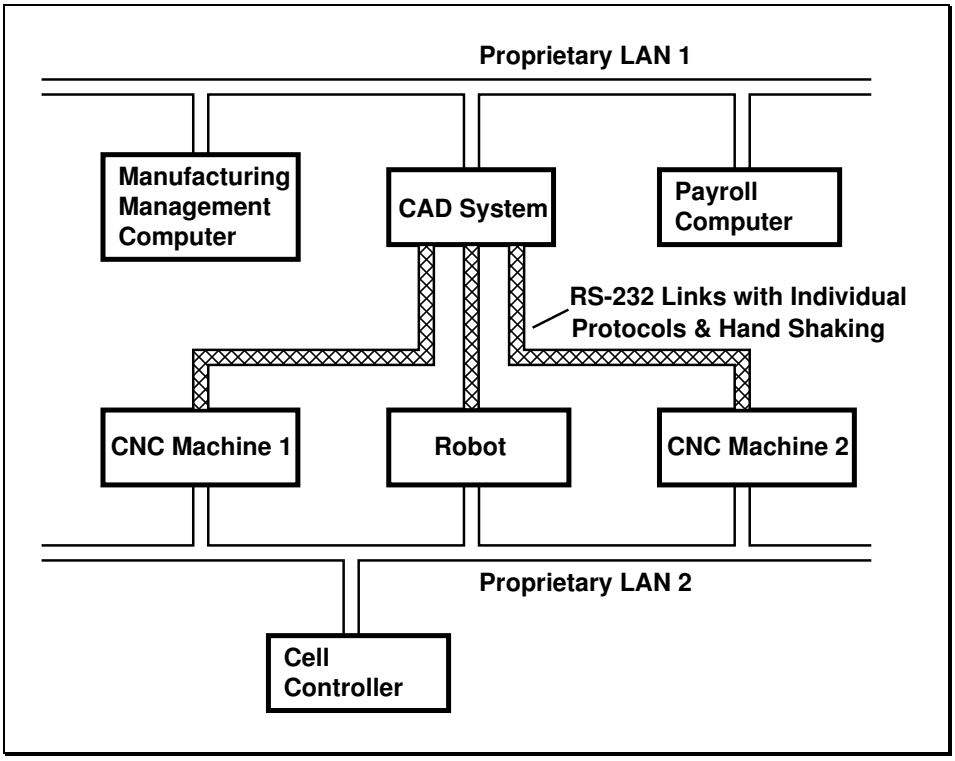


Figure 7.13 - Realities of 1980s CIM

To make matters even worse, bridging between different, proprietary networks is even more difficult than interfacing any single device to a network. To emphasise the enormity of the total networking problem, it is perhaps best to recall how a large manufacturing company viewed the situation from their own perspective in the 1980s.

Mike Kaminsky, of the General Motors' MAP task force, which was established to standardise industrial communications, made the following observation of the networking problem (as it existed at General Motors) in a 1986 issue of the IEEE Spectrum magazine:

> *"Only 15 percent of the 40000 programmable tools, instruments, controls and systems already installed at General Motors facilities are able to communicate with one another. When such communication does occur, it is costly, accounting for up to 50 percent of the total expense of automation because of the wiring and the custom hardware and software interfaces needed".*

We shall examine the role of proprietary networks and major attempts at industrial networking standards (such as MAP) in the next chapter.