INTRODUCTION TO CONTROL NETWORKS IN AN INDUSTRIAL SETTING
PDHengineer.com Course No. IC-3001

Chet S. Barton, P.E.

1.0 Introduction

Communication networks in the industrial arena have, in the past decade, revolutionized the way facilities are controlled. They have made centralized control centers possible, with a wider range of features and more flexibility than ever before. High data transfer rates have allowed for more efficient data storage, trending, alarming, and analysis. The drawbacks that plagued the early generations of networks have been solved, for the most part, making them reliable enough to be used in the most critical of applications.

We take communication for granted, and while we practice it on a routine basis, it can be complex to achieve on an artificial level. We'll start with an example of two people wanting to communicate: person one has an idea he wants to share with his friend, person two. He must first formulate his idea into a coherent thought. He must then organize that thought into an expression that makes sense. He must structure that expression into a sentence that will effectively convey his idea. At this point he has a message that he must transmit to the other person. At the appropriate time, he begins vibrating air with exhaled breath across his vocal cords, and modulates those vibrations with his mouth. The message is transmitted, and person two must receive the message. Person two must recognize that person one is about to transmit a message, and prepare to receive it. He must perceive the vibrations in the air with his eardrum, and convert the vibrations into electrical stimuli to his brain, where it is transformed into a sentence in his conscious mind. He must then decode the sentence to determine its literal translation, do error checking, and then extract the real, intended meaning from it.

This sounds cumbersome, but think in more everyday terms: you're at work, and getting hungry. You notice it is nearly your normal lunch time. You walk over to your friend's desk. When he notices you walking toward him, he looks up. You say "Are you ready for lunch?" He nods, gets up from his desk, and proceeds to the door. You did all of the previously mentioned tasks with hardly a thought.

The definition of a network is two or more devices connected by some means so they can share information. The "means" is what we will address here. Additionally, even though the most general interpretation of the definition could include many manifestations (including our two friends going to lunch), we will focus on data communication between devices commonly found in industry. If the problem is broken into manageable parts, we can deal with each one effectively.

1.1 Communication Nomenclature

There are several techniques used to transmit information. Nearly all data network systems in use today, use binary digits (bits), a series of 1s and 0s, to send information, but there also must be methods of carrying the bits across the network.

Baseband involves the use of the entire bandwidth of a channel to transmit a single signal, using one carrier frequency. **Broadband** divides multiple analog signals into different frequencies and transmits them simultaneously (multi-mode fiber optic signals are one example of this).   Baseband hardware and associated cable are typically less expensive. Broadband requires high quality coaxial cable, which has the disadvantage of being expensive, as well as heavy, stiff, and difficult to work with. **Bandwidth** is the range of frequencies that a given carrier is able to effectively transmit. The rate at which data can be sent depends on the bandwidth of the cable, and is expressed in “baud”. **Baud** is actually the rate of signaling events (changes in frequency, amplitude, etc.) **Bits per second** is not necessarily the same as the baud rate and should not be confused, even though the terms are routinely used interchangeably. For instance, if the modulation allows for four distinct states rather than the traditional two (1 & 0), two bits could be conveyed with every state. This would mean that the rate of bits-per-second would be twice as fast as the baud rate. The baud rate is the same as the bits-per-second rate if and only if each signal element is equal to one bit exactly.

Messages are assembled into packets with formatting and addressing information, along with the data. The general form of a message packet or frame is a leading **header** (sometimes called the *preamble*), the **data** area (called the *payload*), and the **trailer**. The header contains addressing and error checking information, the data area contains the actual data being transmitted, and the trailer contains more error checking and message management information (e.g. parity and stop bits). **Parity**, a simple error checking method, uses the number of 1s in a byte (odd or even) to determine if the byte was received correctly.

Simplex transmissions are only in one direction, all of the time. **Half-duplex** is bidirectional communication allowed in one direction at any given time, and **full-duplex** is bidirectional transmission in both directions simultaneously. In addition to this, **synchronous** (clocked) transmissions are timed so that both devices know exactly when a transmission will begin and end, whereas **asynchronous** (un-clocked) transmissions must mark the beginning and end of messages. Synchronous transmission is usually faster than asynchronous, but the timing issue between two remote machines can introduce problems causing asynchronous transmission to be simpler and less expensive, and therefore more widely used. Asynchronous transmission does, however, introduce extra control bits into a message, which slows the rate that actual data can be transferred.

1.2 The OSI Seven-Layer Model

The OSI (Open System Interconnection) Reference Model (developed by the International Standards Organization, or ISO, so don't get confused) has been developed to define the communication process. It contains seven layers, and is independent of technology:

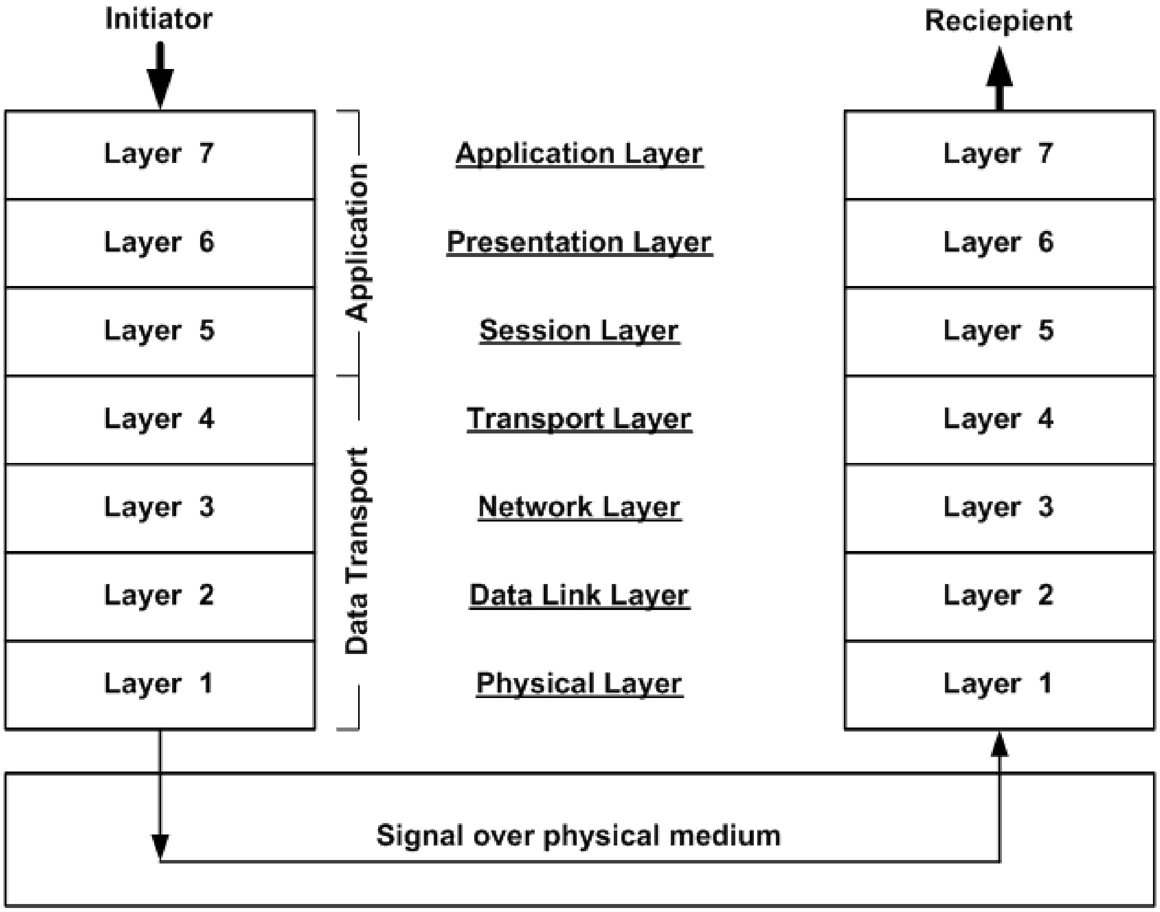


Figure 1 – The Seven Layer OSI Model

Layer 1 - **Physical** - Defines the media characteristics (e.g. cable, connector, voltage levels, modulation techniques, etc.) necessary to originate, maintain, and terminate the link between physical devices.

Layer 2 - **Data link** - Provides the physical addressing, network topology, data packets assembly, managing links between nodes, as well as error detection and correction at the bit level.

Layer 3 - **Network** - Establishes connectivity and path selection between two end systems with logical addressing. It is the layer at which routing across sub-nets occurs.

Layer 4 - **Transport** - Responsible for reliable network communication between end nodes, mid-level control of message delivery, including handshaking, message-level error detection and retransmission.

Layer 5 - **Session** - Establishes, manages, and terminates sessions between applications and manages data exchange between presentation layer entities, the higher level addressing of messages, as well as system control that controls communication sequencing and timing.

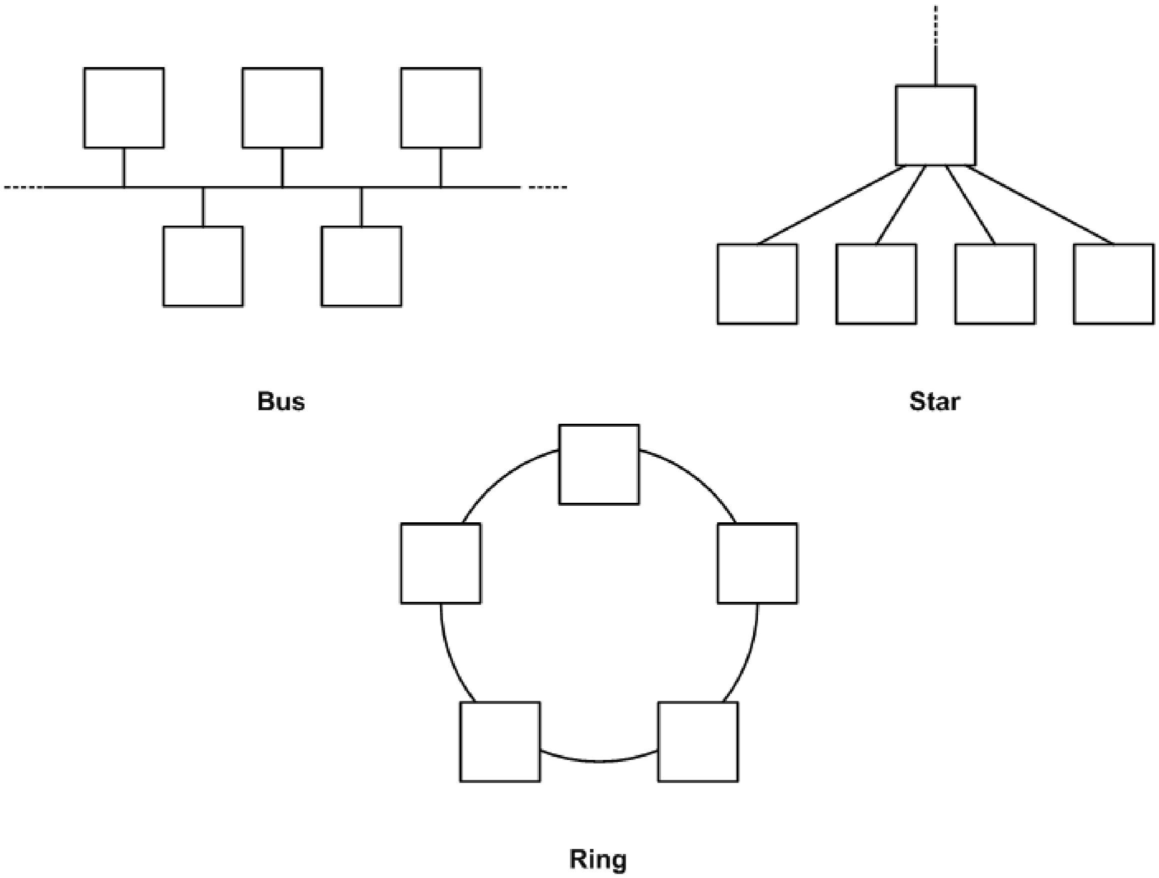
Layer 6 - **Presentation** - Ensures that information sent by the application layer of one system will be readable by the application layer of another by translating the message into the proper format. It is concerned with the data structures used by programs.

Layer 7 - **Application** - Provides services to application processes, the user programs that make a transmission request (such as e-mail, file transfer, and terminal emulation) that are outside of the OSI model. The application layer identifies and establishes the availability of intended communication partners (and the resources required to connect with them), synchronizes cooperating applications, and establishes agreement on procedures for error recovery and control of data integrity.

This discussion will primarily include the transport portion of the model (Layers 1 through 4). Those layers include most of what people mean when they speak of communication networks. The upper layers, referred to as the application portion, have more to do with the programs that manipulate data.

2.0 Network Topology

The arrangement of the devices on a network (sometimes called nodes), is called the topology. There are several that are commonly used, including the following:



Bus

Star

Ring

Figure 2 – Common Topologies

Combinations of the basic types are sometimes used, depending on the application.

It is important to remember that the model is “technology independent”, and is not limited to wire or even a physical connection.

3.0 Physical Media

There are several types of media used in industry: coax, twisted pair, and optical fiber are common examples.

Coaxial Cable (Coax) - A type of cable with a single solid conductor surrounded by a tubular shield, or outer conductor, separated by insulation and oriented about a common axis (hence the name). The shield reduces electrical interference (EMI : Electro-Magnetic Interference & RFI : Radio Frequency Interference). The shield also acts as the signal reference conductor.

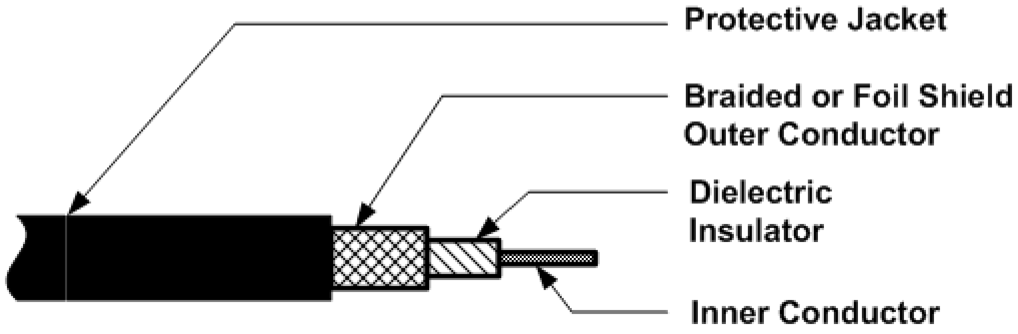


Figure 3 - Coaxial Cable

There are several variations of coaxial cable that are commonly used in industrial networks, designated by an "RG" number. "RG" is a U.S. Army descriptor (U.S. military standard MIL-C-17) which stands for "Radio Guide". Coax usually supports radio frequencies (RF signals) from 50 to 500 MHz, and is commonly (although not exclusively) used for broadband signals. In general, the larger diameter coaxial cables offer greater performance due to their more rigorous construction; however, they have more stringent installation requirements.

Cable Type	Characteristic Impedance	Common Usage
RG-6	75	Broadband, Carrier Band (Drop)
RG-8	50	Thick Ethernet
RG-11	75	Broadband, Carrier Band (Trunk)
RG-58	50	Thin Ethernet
RG-59	75	Broadband Drop
RG-62	93	ARCnet

Note: Some references include the dash in RG-X, others do not.

Figure 4 - Characteristics of Coaxial Cables

Characteristic Impedance is one parameter used to describe the attributes of a particular cable. A simplified explanation is that it represents the impedance one would measure if the cable were of infinite length. High frequency signals can reflect from the ends of a cable and disrupt communication, but there can be no reflection if there is no end (i.e. infinite length). If a terminating resistor with a value equal to the characteristic impedance is placed at the end of a cable run, it approximates an infinite cable, and minimizes signal reflections.

Shielded Twisted Pair (STP) - Cable with a number of individually insulated conductors, twisted into pairs, and surrounded by a shield to reduce EMI/RFI. The pairs may be individually shielded or have one overall shield, or both. STP is more expensive than UTP, but is more resistant to interference, making it the cable of choice for most industrial networks.

Unshielded Twisted Pair (UTP) - Cable very similar to STP, except without any shielding. It is very commonly used in informational Ethernet applications because of its less expensive construction and ease of use. A system has been devised by the EIA/TIA (EIA/TIA-568 standard) to define physical specifications and rate performance of cables for different applications.

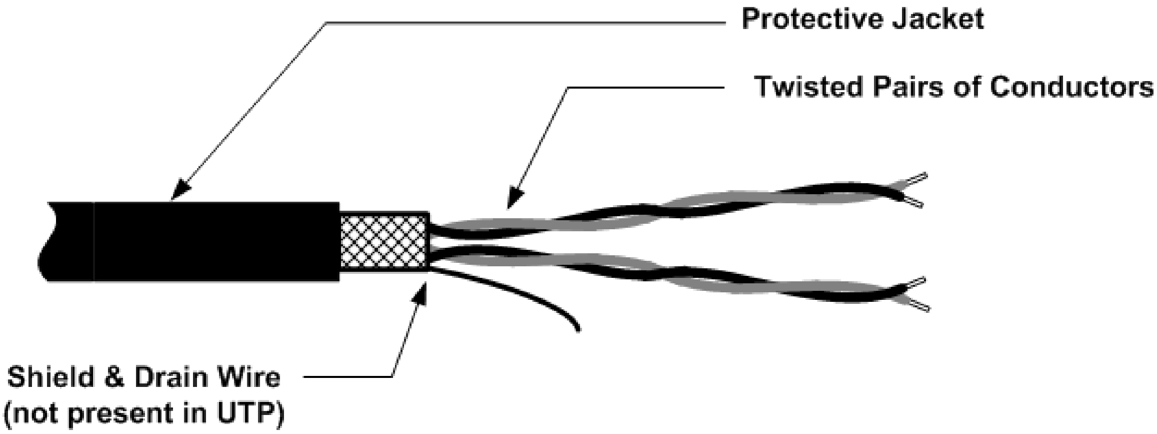


Figure 5 – STP /UTP Cable

Cable Type	Data Rate	Common Usage
Cat 1 (Category 1)	N/A	Voice Grade Analog (Not suitable for data)
Cat 2	4 Mbps	Digital Voice
Cat 3	10 Mbps	10BaseT
Cat 4	16 Mbps	Token ring
Cat 5	100 Mbps	100BaseT
Cat 6 (5e)	1000 Mbps	1000BaseT

Figure 6 – Characteristics of UTP Cable

Category ratings are determined by physical factors such as the number of twists per foot, characteristic impedance, etc. Installation of cables should be made with cable construction in mind to maintain these qualities. More on installation practices will be covered later. Bulk cable is typically constructed of solid conductors, which performs better over long distances, while patch cables are usually stranded conductors to offer more flexibility inside cabinets.

Optical Fiber Cable

Optical fiber has the ability to transmit higher data rates for longer distances than copper media; however, it costs much more and is usually more difficult to work with. Fiber optic cables have the added important benefit of being completely immune to EMI and RFI. Optical fiber systems are non-spark-producing so they can more easily be used in explosive areas. The fibers are roughly the size of a human hair and are insulated with a buffer. There are typically several fibers, all surrounded by an outer jacket in a cable. Because of the size of the actual fibers, they make up only a small part of the overall cable size. Most of the cable is buffer, filler, and strength materials. This makes most fiber cables roughly the same size (about 0.5 inches, or 1.27 cm in diameter), regardless of how many strands are in the cable; size of the cable has more to do with construction and jacket type.

There are two types of optical fiber: **Single-mode** and **Multi-mode**. Single-mode fiber is smaller in diameter (usually 8.3 micrometer core) than multi-mode fiber (usually 62.5 micrometer core). Both usually have a 125 micrometer cladding. Single-mode fiber is meant to accommodate a single (frequency) signal, and so can transmit a higher power signal. Multi-mode fiber can handle multiple frequencies at once, but the signals can interfere with one another if their power level is too great. This means that a single-mode signal can be sent a greater distance than a multi-mode signal. There are, however, trade-offs: the single-mode signal is typically launched with laser equipment, which is more expensive, while multi-mode signals are launched with less expensive LEDs operating around the infrared wavelength. Frequencies of light from an LED (Light Emitting Diode) or Laser (Light Amplification from the Stimulated Emission of Radiation), usually from a laser diode, are inserted into a shaft of plastic or glass (glass is more common) with a particular index of refraction. This shaft is surrounded by another, outer shaft of material with a different index. The difference in indices causes the light to reflect back toward the center of the inner shaft. Single-mode optical applications typically operate in the 1300/1550 nm (nanometers) range, where multi-mode equipment normally operates around 850/1300 nm. The visible portion of the spectrum is in the 400 to 700 nm range.

***Important Safety Note** :* Even though the lasers and LED's used in fiber optic communications operate at a much lower power level than the ones ordinarily used in other industrial or medical applications, they can still be powerful enough to harm sensitive eye tissues. NEVER look directly into the end of a fiber or port unless you make absolutely certain that there is no power on the circuit! The wavelength of light used is invisible to the unaided eye and may not be detected until damage is done.

There are also two types of fiber optic cable construction: **tight-buffered** and **loose-tube**. Loose-tube cables are filled with a gel, which insulates and protects the fibers but allows movement within the jacket so that vibration, thermal expansion and contraction, etc., do not inflict catastrophic damage on the delicate fibers. Tight-buffered cable is smaller in physical diameter, more flexible, easier to work with and terminate, and generally more suited to indoor applications than loose-tube.

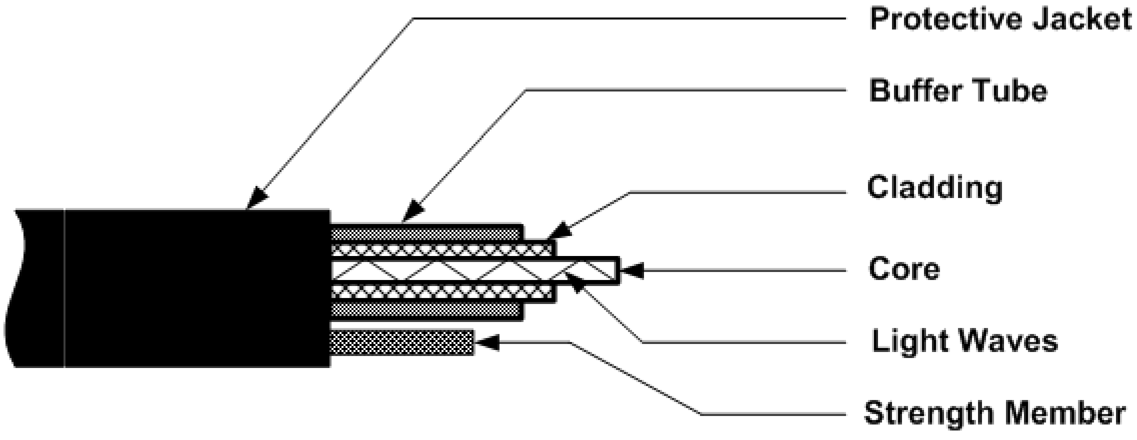


Figure 7 – Optical Fiber Cable

4.0 Arbitration of Physical Media

Physical arbitration has to do with determining which device has the right to speak, or to control the network. Some methods are:

Token Passing - A method of media access control in which devices take turns using the network, by passing a special data packet, known as a **token**. When a device has possession of the token, it may access the network for a specific period of time before it must pass the token on to the next device.

Multiple Random Access - A method of media access control, in which all devices simultaneously have the same priority and negotiate for control of the media. **CSMA/CD** (Carrier Sense, Multiple Access/Collision Detect) is the most common form of MRA methods. “Multiple access” refers to the many stations that share equal access to a common media. “Carrier sense” means that all devices listen to the media and will only claim control if no other devices are currently active. A device may only retain control for a limited amount of time before relinquishing it to give other stations a chance. The “collision detect” feature is a method of arbitrating the inevitable event where two stations attempt to claim control simultaneously. Upon hearing the resulting corrupted signals (a collision), both stations stop and wait for an arbitrary time before trying again.

5.0 Logical Schemes & Arbitration

Client / Server & Polling (Master / Slave) - In a client/server arrangement, a client makes a request of the server, which responds to the client. Servers may not speak until spoken to. There may be many clients and many servers, however, if more than one client is present, another arbitration method (usually physical) is required among them. Roles of devices may change, and a node that at one point is a client making a request of another, may later be a server to that very node. **Polling** is a similar access method in which one device makes a request from another, but the connotation is that of a more regular, timed interval than that of client/server. The master/slave arrangement is similar in nature, in that the master makes requests of the slave. The difference is that the roles do not change: masters remain masters and slave devices remain slaves. (There is a “floating master” technique, which is very similar to client/server.)

Peer-to-Peer - In a peer-to-peer network, no one device is the controlling authority. Each is equal in priority; that is to say, they are peers. Devices take turns writing to, or making requests of one another.

Publisher / Subscriber - Basically the reverse of Client/Server. This method has one or more suppliers of information, which broadcast information to all other nodes, without waiting to be polled. Those nodes which need the information receive it, and those that do not, ignore it.

6.0 Open and Proprietary Protocols

***Open Protocols* -** The term “open protocol” literally refers to a communication standard that is available for use by the general public. The implication is that it is not developed by a particular company, but by an industry cooperative. In reality, many protocol standards have been developed by companies and later released, to be adopted by many manufacturers. Technically, there are few truly open systems; mostly there are degrees of openness. Not any one standard is wholeheartedly embraced by every manufacturer. There are, however, some standards which are common and pervasive throughout the industry. Modbus, for instance, was developed by the Modicon company for proprietary use, but the standard is freely distributed and has become an industry defacto “open” standard. The standard is, however, still controlled by Modicon. Systems using open protocols usually cost less to install and modify than those using proprietary standards.

***Proprietary Protocols* -** Proprietary protocols are those developed and maintained by a particular company. Equipment is typically limited to a few manufacturers. The advantages of a proprietary protocol have mostly to do with strict compatibility between components, giving ease of use and increased performance. The disadvantages are that the user is restricted to the offerings of a few vendors, with limited flexibility and greater general cost.

6.1 Open Industry Standards

The EIA (Electronic Industries Association) has developed several “Recommended Standards” (RS-XXX) to aid in the ease of connection. They are physical standards that specify cable, connectors and electrical properties.

RS-232C - Probably the most widely used and versatile physical standard. It was developed mainly for the interface between DCEs (Data Communication Equipment, typically a modem, is the interface which actually performs the business of communicating), and DTEs (Data Terminal Equipment, the device ultimately attempting to communicate) in public telephone network services. Because RS-232C is an unbalanced system, the distance between devices is limited to roughly 15 meters (45 ft., at 19.2 Kbps), which is its largest downfall, but the inherent flexibility makes it a popular choice for many manufacturers. Unbalanced systems modulate a voltage signal on one data line, with respect to a signal ground. Even though it is possible to successfully transmit data at a slower rate for longer distances (maximum of 100 feet, or 30.3 meters), the standard recommends that the cable be no longer than 50 feet (15.2 meters). The EIA standard allows both synchronous and asynchronous transmissions. It also permits a range of data rates up to 19.2 Kbps, varying data lengths and bit codes (parity, stop bits, etc.), which are normally selectable in the port configuration. The various modes (simplex, half/full duplex) are also available. This flexibility, while convenient, allows for a wide variety in interpretations, and can lead to some incompatibility between manufacturers (ironically, this is contrary to the concept of a standard, which is a disadvantage of the "open" model). RS-232C was originally designed to use a DB-25 connector, but in practice most of the handshaking and timing capabilities are rarely used, so a DB-9 connector can be substituted at the user's discretion. The RS-232C interface is usually fashioned with a female connector on the DCE end, and with a male connector on the DTE end. The most common use involves six data lines, along with two ground connections:

Protective Ground - May be connected to an outer screening shield conductor, and tied to the equipment chassis. (Only on one end to prevent ground loop problems)
TD - Serial "Transmitted Data," an output.
RD - Serial "Received Data," an input.
RTS - "Request to Send," an output.
CTS - "Clear to Send," an input.
(RTS and CTS manage speaking and listening functions in a half-duplex arrangement.)
DSR - "Data Set Ready," an input.
DTR - "Data Terminal Ready," an output.
(DSR and DTR give an indication that the device is powered and ready to communicate.)
Signal Ground - Common signal reference; should not normally be connected to chassis ground.

Signal levels which differentiate between a one or a zero assume two, bipolar voltages. Standard circuitry uses a range of from +3 to +25 volts DC to represent a zero, and a range of from –3 to –25 volts DC to represent a one. The relatively large voltage levels used helps make the RS-232C link more resistant to EMI/RFI noise, but with a common signal ground, there is no double-ended signaling (i.e. unbalanced), which allows common-mode noise to exist.

RS-422 - Being a balanced system, RS-422 makes longer distances and faster transmission rates than RS-232C possible. RS-422 can have multiple receivers, but only one line driver per twisted pair of wires. To facilitate full-duplex operation, two separate channels are used that would be two completely separate RS-422 links except for the fact that they reside on the same device(s). Because it is a balanced, or *differentiated* system, it is very resistant to EMI/RFI. A balanced system uses two data lines and modulates a signal voltage between the two. It is capable of 10 Mbps at distances of 4000 feet (1219 meters). The RS-422 standard recommends a 24 AWG twisted pair cable with a 100 ohm characteristic impedance. Cat 5 cables, used in Ethernet, meet these requirements and is widely available, making it a good choice for RS-422 installations. In keeping with the practice of terminating the cable with the equivalent of the characteristic impedance, a 100 ohm resister should be connected across the “A” and “B” lines.

RS-485 - A serial interface standard which (like RS-422) is a balanced system, except that it uses a tri-state line driver. The third, high-impedance off-state, allows an inactive device to sit quietly on the network, making multiple drops (up to 32 bi-directional line drivers) easier to manifest. When a device is not transmitting, its line-driver must go into the high impedance state, so as to not interfere with any other transmission. It also permits even faster rates over longer distances--100 Kbps at up to 4000 feet (1219 meters). It can operate on one pair of twisted pair in half-duplex mode, or on two pairs in full-duplex mode. The full-duplex mode is limited, with only a master/slave arrangement is allowed. Unlike 422, an “enable” function must be used. Exactly how this is accomplished varies between manufacturers. The two-wire/four-wire distinction can be misleading, as a ground wire is also required. This would appear to blur the distinction between balanced and unbalanced systems, except that the signals are not referenced to the ground in a balanced system. The two-wire system is less expensive to implement than the four-wire version. Even though the 485 standard does not specify a cable, the same twisted pair wiring used in 422 applications can be used to implement 485 as well. The two wire system is slower because *bi-directional* transmissions occur over the same channel, and the line driver must be put into the high impedance state when it is finished before any other driver can begin.

6.2 Open Protocols

Ethernet – probably the most widely used network protocol, it is mature and highly developed. It is mainly found in corporate settings, but is rapidly finding acceptance in control applications. It was developed by the Xerox corporation, but hardware is widely available from a multitude of vendors. It is very similar to the IEEE 802.3 standard, but is technically not the same, although the term “Ethernet” is loosely used to describe both. Both are serial, broadcast type networks, with the main difference being that 802.3 specifies several different physical layers, whereas Ethernet defines only one.

Ethernet was originally meant for the office environment, so connectors and other equipment seldom met industrial standards. This is changing as traditional manufacturers of industrial networking equipment develop hardened components that can withstand harsh conditions. Since it uses the CSMA/CD media access method, earlier implementations of Ethernet had the added disadvantage of being rather **indeterminate** (i.e. non-predictable), which was not a problem in most office settings, but is not tolerable in most control applications. This problem was often the result of poorly designed network arrangements, due to the lack of the proper equipment. Office network traffic tends to be sporadic, which minimizes problems with message collisions, coupled with the fact that these messages are not normally time sensitive. This is the exact opposite of most control strategies which exchange data fairly constantly and require that values be refreshed regularly. Adding more devices only serve to increase traffic, causing more collisions. For these reasons, Ethernet was thought to be inappropriate for industrial control use.

The advent of switches and routers has alleviated most of these problems by separating collision domains, minimizing collisions and facilitating more timely message delivery. For all of these reasons, Ethernet is poised to become one of the most popular industrial networking protocols, as most major manufacturers are introducing industrially hardened Ethernet interfaces, switches and connectors, as well as industrial control message protocols (Modbus/TCP, Ethernet/IP, Profinet, Fieldbus HSE, to name a few).

Ethernet utilizes a baseband signal. The most common configuration is over twisted pair wires in either a bus or star topology, although optical fiber is becoming more prevalent. A notation for describing Ethernet signal types is: N <Signal-Type> X, where N represents the signal rate in megabits per second (Mbps), the signal type is either baseband or broadband, and the X represents a special characteristic, such as “5” meaning the maximum distance is 500 meters, or “T” meaning the signal is transmitted over twisted pair, or “F” meaning it is transmitted over optical fiber. A designation using “Base” as the signal type (e.g. 10Base2) is not referring to the base 2 numbering system.

Signal Name	A.K.A.	Data Rate	Max. Distance	Special Designation	Media
10Base2	Thinnet	10 Mbps	185 Meters (606.8 ft)	Approx. 200 m.	50 Ω Thin (RG-58) Coax
10Base5	Thicknet	10 Mbps	500 Meters (1640 ft)	Approx. 500 m.	50 Ω Thick (RG-8) Coax
10BaseT		10 Mbps	100 Meters (328 ft)	Twisted Pair Cable	Cat. 3 or 4 Twisted Pairs
10BaseF		10 Mbps	2000 Meters (6561 ft) (1.2 miles)	Fiber Optic Cable	
100BaseT	Fast Ethernet	100 Mbps	100 Meters (328 ft)	Twisted Pair Cable	Cat. 5 Twisted Pairs
100BaseF	Fast Ethernet	100 Mbps	400 Meters (1312 ft)	Fiber Optic Cable	
1000BaseT	Gigabit Ethernet	1000 Mbps or 1 Gbps	100 Meters (328 ft)	Twisted Pair Cable	Cat. 5e or 6 Twisted Pairs
1000BaseF	Gigabit Ethernet	1000 Mbps or 1 Gbps	220 Meters (722 ft)	Fiber Optic Cable	

Figure 8 - Common Ethernet Signal Types

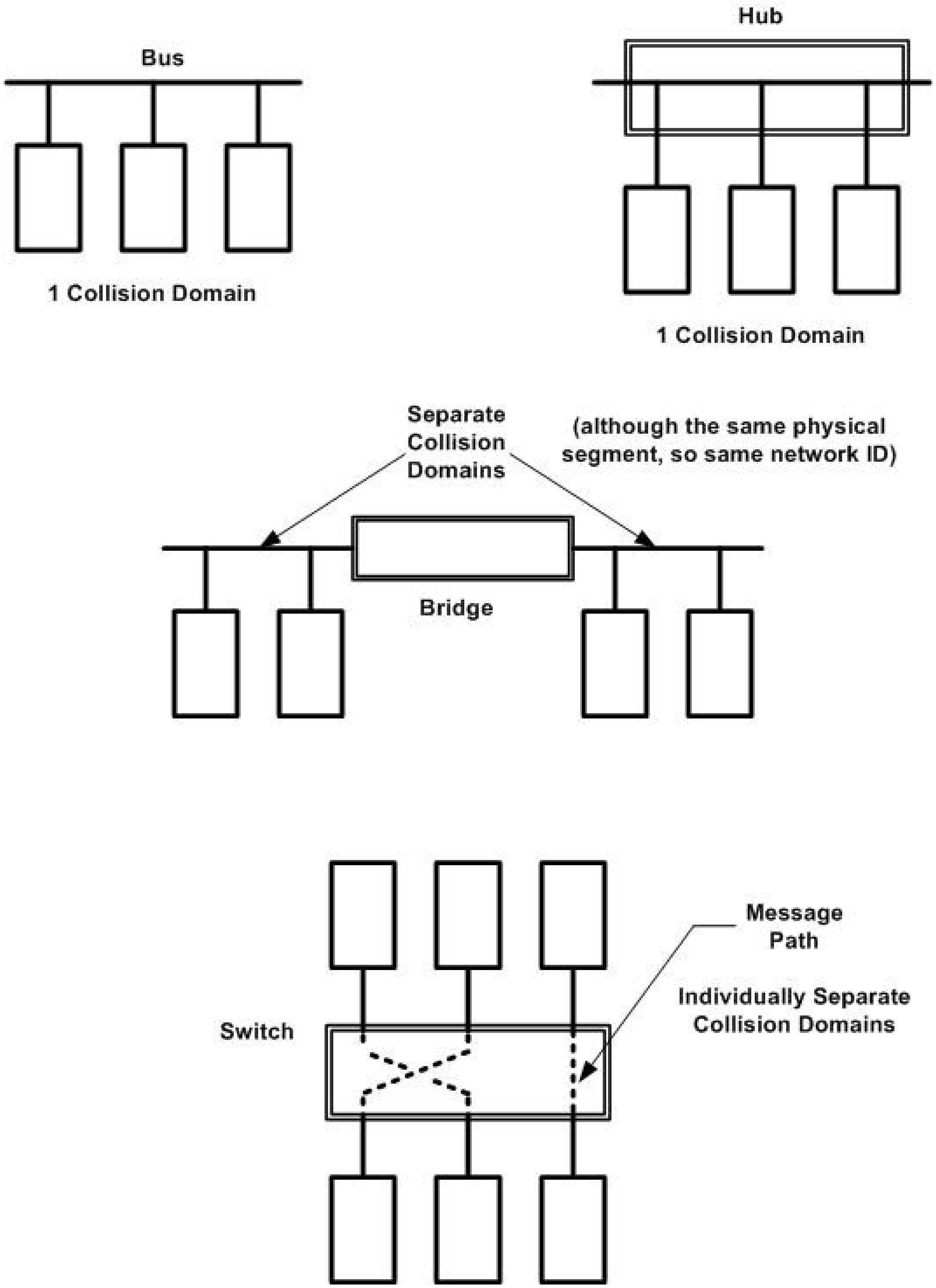


Figure 9 - Collision Domains

The most common protocol used with Ethernet is **TCP/IP**, and this is the route that the new industrially hardened Ethernet implementations are following. TCP (Transmission Control Protocol) defines the source port and destination port numbers that allow data to be sent back and forth to the correct application running on each device. The TCP portion of the message also includes error checking and other housekeeping information. It operates on the Transport layer of the OSI model. IP (Internet Protocol) is responsible for inserting routing information into the header of each message on the Network Layer. It is said to be “connectionless” because it does not concern itself with any handshaking practices. When IP does not know the path to a destination, it forwards the packet to the host's default gateway. The default gateway is then responsible for routing the packet to its destination. That default gateway might, in turn, pass the packet to its own default gateway if it does not know the path.

The IP address of a device can be **statically** or **dynamically** assigned. A statically assigned address is one that is manually typed into the device’s configuration, and does not change. A dynamically assigned address is one that is assigned by a network master when the device boots up and joins the network environment, and is generally different every time. Dynamically assigned addresses make corporate networks easier to manage, but are not common in control networks.

An I/P address consists of 32 bits, and is split into four groups of eight bits (called octets), separated by periods whenever it is written or typed in, to make it easier for people to deal with. The first part gives the network ID number, and the second part gives the particular device ID number, which must be unique within that network segment. All devices on a particular physical segment share a common network ID. Which bits correspond with which part depend on the class of address and the subnet mask. The subnet mask has a structure similar to the IP address, with the ones indicating the network ID portion of the IP address, and the zeros indicating the device number.

Modbus - Another type of open protocol standard, which was originally developed by a private company. It was introduced in 1979, as one of the first industrial network standards. It has been so widely accepted that it is a defacto industry standard, even though the standard is still controlled by Modicon. It was developed by Modicon as a structure for data representation, and is independent of the physical layer. It can be implemented over any transmission medium, but is most commonly seen used with RS-422, RS-485, and most often with RS-232. It is a serial transmission technique which uses master-slave arbitration. A master can communicate half-duplex style with up to 247 slaves. It can make broadcasts to all slave devices, but in that case, slaves are not allowed to send a response.

Modbus can operate in two modes: ASCII and RTU. The mode is selected along with the other serial port communication parameters (baud rate, parity, etc.). It simply defines how data is packed into the message frame. In ASCII, eight bits of data are sent as two ASCII characters. This allows for more reliable transmission, with gaps of up to one full second to elapse without causing errors. RTU mode packs two four-bit hexadecimal characters into the message data area. RTU offers greater data density, and therefore higher throughput for a given baud rate, so it should be used whenever possible (which is nearly every case). The Modbus packet is included in the data area of a standard message packet. A newer version, Modbus/TCP has been issued to ride on the Ethernet specification. The Modbus message packet is inserted into the data area of the Ethernet message packet.

Modbus can typically transfer 300 registers per second at 9.6 Kb, with the maximum speed being 19.2 Kb. While not as fast as most other industrial standards, but its simplicity and availability make it very popular with many manufacturers.

6.3 Proprietary Protocols

Modbus Plus - an RS-485 based, peer-to-peer network protocol, which uses the Modbus data structure. It transmits at a rate of one megabaud. It allows up to 32 nodes on a segment of up to 1500 feet (455 meters). A repeater allows an additional segment to be attached for a maximum of 64 nodes, over 3000 feet (909 meters). Another repeater may be used to increase the maximum distance to 6000 feet (1818.2 meters), but the 64 node limit remains. This distance restriction is for the recommended shielded twisted pair cable, but the maximum distance using optical fiber is several miles, depending on network and repeater configuration. Modbus plus uses the bus physical topology, with a token passing arbitration method. It is capable of transferring 20,000 registers per second. Devices on a MB+ network are given network address, and the device with the lowest address assumes the role of network master. After a node uses its allotted time, it releases the token and passes it to the next higher addressed device. Multiple network segments can be joined together, and Modbus Plus allows routing up to five layers. By using appropriate modules with dual ports, the network can use redundant cabling.

The old style of connectors went directly on the trunk cable in a daisy-chain fashion. A new style has been introduced with taps and drop cables. These make for a neater installation, and the taps can be configured as in-line or terminating with jumpers. It is very important to adhere to the published guidelines that require a minimum of 10 feet (3 meters) between taps to prevent reflected and corrupted signals. Both the old and new versions incorporate a 120 ohm terminating resister on both ends.

DataHighway Plus - The Allen-Bradley/Rockwell Automation company has developed a similar proprietary network communication system that has become so common and popular, it is pseudo-open. As in previous examples, DH+ is proprietary, but many manufacturers produce compatible equipment which fill a variety of requirements. It is token passing, operates at 57.6 Kbps, and also uses a pair of twinaxial wires. DH+ specifies Belden 9463, which is commonly known as “Blue Hose.” The network can be arranged in a daisy-chain or bus (trunk/drop) configuration, which uses taps with BNC connectors. There are 64 nodes allowed, and the trunk length depends on the configuration, but the maximum is 10,000 feet (3050 meters). The trunk should be terminated with a 150 ohm resister at each end.

7.0 Installation and Troubleshooting Topics

As seen in the previous discussions, communication networks can be very complex, and can have exacting requirements. Problems can easily occur, but most of these systems have evolved over many years in many installations, and will perform satisfactorily if installed and maintained properly. The following section offers advice on how to avoid and/or correct any of these problems.

Aside from design-type installation considerations, care needs to be given to physical installation of network systems, especially cabling. The cables used in modern communication systems have very particular construction specifications, which if not maintained, will degrade the performance of the entire system. Factors such as cable stretch during installation and use, bending radius, and termination techniques can alter cable construction. It has been said that most network problems are due to cable issues.

Preventing Problems - During the discussion of the various standards, protocols, cables and connectors, many specifications and parameters are outlined. These qualities must be maintained to ensure continued performance. Careful attention to installation instructions and general guidelines will go a long way to minimizing troubleshooting later:

1. When installing cables, maintain structural integrity by not pinching, kinking, stretching, or abrading it.

2. When terminating conductors and connectors, manufacturer’s guidelines offer criteria for preserving the ratings for speed and noise resistance. For instance, untwisting too much of a twisted pair cable at the termination will degrade the performance of the whole system. A Category 5 cable must have no more than 0.5 inches (1.27 cm) untwisted at the end, or it will no longer meet Category 5 requirements, and may not support a 100 Mbps signal. Many topologies have segment length requirements (both maximum and minimum) which are meant to reduce signal reflections and corruption.

Most problems this author has encountered has come from a single strand of a wire hanging out of its allotted terminal and touching another or shorting to ground.

3. Be sure segments are terminated properly (with terminating resistors and/or capacitors), according to instructions to prevent reflections and corrupted data.

4. Keep signal wires away from sources of EMI and RFI. If communication cables must be run near power wiring, be sure to cross at a 90 degree angle to reduce induced voltages. While shielding can reduce the effects of EMI/RFI, improper grounding, particularly grounding at both ends of a long run, can allow ground loops to form and actually introduce interference.

5. When installing interface cards and other network equipment, be sure that all hardware and software configuration settings are correct and consistent (speed, parity, etc.). Be sure that settings for cards installed into PCs don't interfere with other devices. NICs require configuration, including hardware interrupt (IRQ) and memory locations, which if in conflict with another device, will cause malfunctions.

Diagnosing Problems - The first rule of troubleshooting is to check the obvious first:

1. Check to be sure that all connections are made securely.

2. Make sure all components are seeing the network by observing LEDs. Most interfaces will have status lights that will flash a particular code to indicate errors.

3. Make sure communication is enabled or being initiated by some node.

If devices are still not able to establish a link, more aggressive methods are in order:

4. Completely disassemble the network, and rebuild it one piece at a time (segment, device, etc.). Try to establish communication between only two nodes and then add other components slowly.

5. Many manufacturers will supply diagnostic software with equipment or on web sites. These may be used to monitor port activity and diagnose potential errors.

6. Check each connector for continuity (with all connectors unplugged from devices). Remember that the resistance read should be that of the equivalent circuit of the cable, including cable resistance (typical value is 1 ohm per 100 feet, or 30 meters), as well as that of any terminating resistors (in parallel, so R/2). With the terminators removed, there should be an open circuit between signal conductors.

8.0 References

1. Copyright (2002) from Instrument Engineer's Handbook by Chet Barton/Bela Liptak. Reproduced by permission of Routledge/Taylor & Francis Group, LLC

2. Real-Time Control Networks; ISA, Miklovik, 1993