Volume 1 • Issue 9
November–December 2000

Introduction to
Industrial Ethernet, Part 6.

Part 5 was featured in Issue 8, SEPT. - OCT. 2000. If you would like a copy, please send your request to **info@ccontrols.com**

the EXTENSION

A Technical Supplement to control NETWORK

INTRODUCTION TO SWITCH TECHNOLOGY

Improving the Performance of Ethernet Networks

By George Thomas,
Contemporary Controls

INTRODUCTION

In a previous article we discussed the use of repeaters in order to increase network diameter. The network diameter of an Ethernet network can be increased using repeaters as long as the network diameter does not exceed the collision domain of Ethernet. All Ethernet nodes must be able to recognize the occurrence of a collision regardless of the physical location of the nodes since the detection of collisions is fundamental in the manner Ethernet arbitrates media access. In this article, the concept of switching will be introduced as an alternative to the deployment of repeaters. Switches can not only increase the overall network diameter, but will improve the performance of Ethernet networks as well.

CLASSIFYING DEVICES

Although from the outside a switching hub looks very much like a repeating hub, they are from different classes of equipment. If you study the OSI Communications Model, you will notice seven distinct layers corresponding to different communication services.

At the lowest layer you have the physical layer which is concerned with the actual signals on the medium that represent data. These signals are called symbols and repeaters or repeating hubs receive these symbols and recondition them when extending networks. References such as 10BASE5 and 10BASE-T are physical layer standards.

Above the physical layer is the data link layer which handles the actual transmission and reception of frames sent and received over the physical layer. Issues such as station addressing (MAC or medium access control), framing of the data and error detection are handled by the data link layer. The IEEE 802.3 standard is basically a data link standard although references to physical layer standards are included as well. Bridges operate at the data link layer. A bridge and a switch are one in the same.

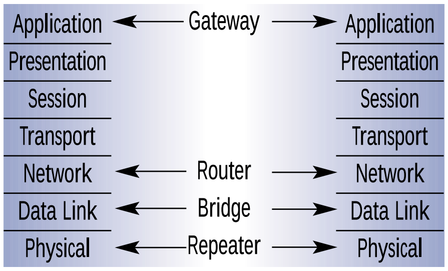

Figure 1–*Repeaters, bridges, routers and gateways operate at different layers of the OSI communications model.*

Above the data link layer is the network layer which addresses the issues of transferring data, not over just one data link, but over multiple data links. This is classified as internetworking with the Internet Protocol (IP) being the most popular internetworking protocol. Routers are used to direct traffic between multiple data links and the transmission units are called packets. Switches do not commonly operate at this layer but there is such a thing as a layer 3 switch. This is actually a router with some switching functionality.

In terms of hardware, the next layer of interest is the application layer seven. This is where the gateways reside when it is necessary to interconnect dissimilar networks and dissimilar protocols. Gateways are aware of the actual application being run while all other devices such as repeaters, bridges and routers are not. We will concentrate only on one class of device called the bridge.

BRIDGE (SWITCH) CONSTRUCTION

A switch is a bridge and the terms will be used interchangeably. The original bridges were two port devices interconnecting two similar data links to form one larger data link. If this can be accomplished without disruption, the bridge is considered a transparent bridge since communication within a data link or between data links appears the same. You may think that we are describing a router but we are not. A router would consider each data link as an actual network with a corresponding network address. A bridge considers each individual data link as part of one larger data link or one network. The concept of network addressing is not used and individual station addresses (MAC addresses) are not duplicated among the various data links. Unlike the traditional bridge with two ports, the switch has several ports and is usually referred to as a switching hub or just a switch.

Unlike a repeating hub, a switch has basically the same Ethernet interface on each of its ports as found on an Ethernet host adapter. That is because each port must function just like another Ethernet device. It must be able to receive and decode Ethernet frames and test for frame integrity as well as assemble and transmit Ethernet frames. However, each port does not necessarily require its own MAC address as would be required by an Ethernet host adapter. Each switch port functions in promiscuous mode by receiving all frames on its port independent of destination MAC address.

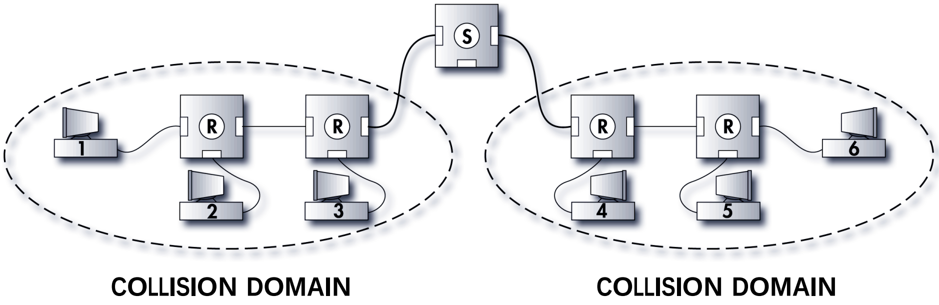

Figure 2–*Collision domains terminate at the switch port.*

During transmissions, the Ethernet port masquerades as the originating device by assuming its source address. Therefore, each port on the switching hub does not require its own MAC address unless bridge addressing is required (the spanning tree algorithm requires bridge addressing).

By having an Ethernet interface on each port, the Ethernet collision domain terminates at the switch port. With a repeating hub, the complete hub is part of the collision domain. By having a switch, the effective network diameter can double with the addition of one switch. This is because the network can be broken into two distinct data links. This is one benefit of switches. The effective network diameter can be increased with the addition of switches. This is especially important at 100Mbps since the collision domain is only 205 meters wide for copper-based systems.

Another difference between a repeating hub and a switch is that the repeating hub must operate at only one speed—either 10Mbps or 100Mbps. A switching hub can have multi-speed ports which can adjust to the capabilities of the device attached to its port. This is called auto-negotiation and different speeds on different ports are allowed. Some switches have fixed low speed ports (10Mbps) and one or more high-speed ports (100Mbps) for connection to servers where most of the traffic will be experienced.

By terminating collision domains at each of its ports, a switch effectively segments the network into separate collision domains. If only one device is attached to a switch port (an Ethernet host adapter or another switch port), this is called microsegmentation. Under these circumstances full-duplex operation is possible yielding no collisions. However, if a shared Ethernet collision domain is present on a switch port (multiple host adapters and a repeating hub), only half-duplex operation is allowed and the switch port must conform to Ethernet's medium arbitration rules.

Switch Operation

To understand the operation of a switch, we will assume that there are no provisions for programming the switch. The switch we will discuss will only modify its operation through a learning process.

Assume our switch has four identical ports. When power is applied to the switch it will behave just like a repeating hub. A data stream received on any one of its ports will be replicated, without modification, onto all other ports except for the arrival port. There will be no subsequent transmission on the port from which the data was received. In this situation the switch is functioning just like a repeating hub. There is, however, one difference. Switches operate at the data link layer and act upon frames. In the general case, a complete Ethernet frame, regardless of length, is received before being transferred to the switch's output ports. Therefore, data latency is introduced that varies with the length of the received frame. A repeating hub operates at the physical layer and acts upon symbols. A received symbol is transferred to the repeating hub's output ports usually within a few bit times. The data latency through a repeating hub is short and independent of the length of the incoming frame.

Let's assume that station A on port 1 is attempting a unicast (one to one) message to station B located at port 2. With a repeating hub or an unlearned switch, all stations on all ports are going to hear this transmission to station B even though they are not part of the conversation. This creates unnecessary traffic on the network that prevents other stations from initiating transmissions since they must defer to this traffic. Only when silence is sensed on the network will a deferring station initiate a transmission. Unlike a repeating hub, a learning switch will note the source address of the transmission from station A on port 1 and will enter into its table the fact that station A resides on port 1. However, at this time the switch does not know where station B resides and, therefore, must send the transmission to all other ports. This is called flooding. Not until station B initiates a transmission will the switch learn that station B resides on port 2. Once station A and B's port assignments are entered into the switch's table, all subsequent unicast transmissions between these two stations will only appear on ports 1 and 2. All other ports will not know a transmission is occurring allowing other stations, not located on ports 1 and 2, to initiate a simultaneous transmission. This is why switches offer improved throughput over repeating hubs.

What happens if station B is physically moved to port 3? If station A again initiates a transmission to station B, the transmission will fail since it will only be delivered to port 2 where the switch thinks station B resides. In order for station B to be found again, it must initiate a transmission. If it does, the switch will note a change in port assignment for station B and change its table accordingly. But what if station B never reports in? Perhaps this station speaks only when spoken to. There is no way for the switch to learn the new location of station B. That is why the switch's table must be aged.

Aging is the process of unlearning. Periodically the switch checks to see if all stations in its table have initiated a transmission within the aging limit. If a

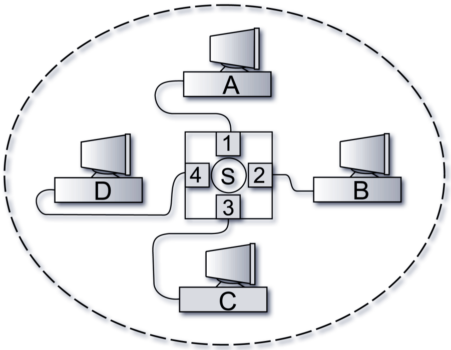

Figure 3–*Switches direct traffic only to necessary ports*

IEEE 802.3 Frame								
56 bits	8 bits	48 bits			48 bits	16 bits	368 to 12000 bits (46 to 1500 bytes)	32 bits
Preamble	SFD	Individual / Group Address Bit	Globally / Locally Administered Address Bit	Destination Address	Source Address	Length	LLC / Data	Frame Check Sequence

Figure 4–*A store-and-forward switch must read in the complete Ethernet frame before forwarding.*

particular station has not, it will be removed from the table. In the example above, station B's entry would be erased. Therefore, when station A initiates a transmission to station B, the switch, finding no entry in its table for station B, will flood all ports allowing station B to hear station A. If station B responds to station A by initiating a transmission, the switch will learn station B's new port assignment and note it in its table. This aging process usually occurs every four to five minutes, so it may take a while to learn locations of devices that are quiet.

Switch Fabric

The beauty of a switch is that, unlike a repeating hub, a switch allows simultaneous transmissions on its ports thereby increasing throughput. This is only true if the switching mechanism within the switch is fast enough to handle simultaneous transmissions on all its ports. If the switch can do this, it is said to be operating at wire speed and it is called a non-blocking switch. If it cannot keep up with the traffic, it may need to queue frames or lose frames. This is called a blocking switch. This switch mechanism within the hub is called its switch fabric and it must be extremely fast for the switch to be effective.

A switch's primary mission is to reliably transfer frames from one port to another. Its secondary mission is to note in its table the port location of various source addresses it learns. Its final mission is to age the table so that stations can be relocated to other ports and still be found by the switch. Depending upon the amount of traffic being handled by the switch, the switch may not be able to do all these tasks with each frame. It is possible that a switch may forgo updating its table when transferring multiple frames meaning that some source addresses will not be noted the first time they appear. The aging process is generally a background process anyway and aging time may vary with traffic.

Data Latency

A repeating hub operates upon symbols while a switch operates upon frames. A switch must receive the complete frame from one of its input ports, observe the destination address, look up the port assignment, note the source address, verify that the frame is not in error and then forward the frame to the indicated port number. This is called store-and-forwarding. At 10Mbps, the longest allowable Ethernet frame will take over 1.2 ms to transfer through the switch. The shortest allowable frame would still take over 500µs to send. The store-and-forward nature of the switch introduces significant data latency. Compare this latency to that of a repeating hub which introduces a delay less than a microsecond. To reduce this latency, the concept of cut-through switches was introduced.

Since the destination address follows the preamble in an Ethernet frame, it only takes about 11µs for the switch to know to which port the frame must be transferred to. The switch could immediately begin transferring the frame to the required port. This, of course, assumes the output port is available. If the port is available, data latency can be reduced significantly. There are, however, problems with this approach.

If the output port is unavailable, the switch would need to queue the frame just like a store-and-forward switch. If the frame was corrupted, as evidenced by a failed FCS test, the switch would have forwarded a defective frame. Defective frames should be discarded by switches and not propagated through the network. However, with a highly reliable local area network, the chance of a failed FCS is rare so this may not be a significant issue. What is significant though, is that runt frames may exist that are the result of collisions. These runt frames are less than 576 bits in length but could be more than the 112 bits of preamble and destination address. Therefore, the switch could be guilty of propagating an error frame by initiating the forwarding of the frame before determining that it is actually a runt. The solution to this problem is the modified cut-through approach where forwarding does not commence until at least 576 bits of frame are received. Only at this time should the forwarding of the frame begin.

Sometimes cut-through operation is not possible anyway. For example, if a switch receives a broadcast, multicast or unknown destination address, it must flood all ports. The probability that all port output queues are simultaneously available for immediate transmission is remote. In this case, the complete frame must be received and sent to the port output queues for eventual transmission.

The significance of data latency can be debated. If each transmitted packet must be acknowledged by the receiving station, then data latency can be important since throughput is impacted by the delay in sending packets and receiving acknowledgements. However, if it is possible to stream the data, the delay in transmission is insignificant since the delay of store-and-forwarding is not accumulative. The delay in sending one frame versus many frames in a row is the same. Streaming of data using the TCP/IP protocol is possible. Knowledge of the transport layer protocol is important when determining if switch data latency is going to be an issue.

Flow Control

With a high-speed switch fabric, there appears to be no bounds to the amount of simultaneous traffic that can be processed by a switch. However, traffic patterns may not be so evenly dispersed. Typically, you will have one port, possibly the port connected to a server or master controller, processing most of the traffic that originate from the other ports. If the switch has no flow control mechanism to limit the traffic being received on input ports and the congested output port has no more buffer available, frames will be simply dropped without any notification. To minimize this possibility, two methods of flow control were developed for switches—backpressure and PAUSE.

Backpressure is used on switch ports connected to half-duplex or shared Ethernet data links. The switch port simply uses the built-in collision detection and backoff algorithm of Ethernet to force collisions on its segment thereby requiring the attached devices to resend their data. When the switch is able to recover, it removes the backpressure.

For full-duplex links there are no collisions, so backpressure will not work. There is instead a PAUSE function developed solely for full-duplex links. A PAUSE frame initiated by a switch port tells the sourcing device to stop sending traffic for a defined amount of time. This scheme only works if the attached device can invoke full-duplex operation and can interpret a PAUSE frame.

IEEE 802.1D.

There is a standard for bridges that is available from the IEEE as standard 802.1D. This standard is entitled, “Information technology—Telecommunications and information exchange between systems—Local and metropolitan area networks—Common specifications—Part 3: Media Access Control (MAC) Bridges.” This standard addresses the uses of bridges and, therefore, switches. There are some interesting parameters in the standard that impact the operation of real-time control networks.

Aging

Aging is the amount of time the bridge waits until it removes a source address from the table due to the fact that the source address has not initiated a transmission within the aging time. The standard allows for an extremely wide range of values from 10 seconds to 1,000,000 seconds. The default, however, is 300 seconds or five minutes, which is what most bridges use.

Bridge Transit Delay

The maximum amount of data latency introduced by a switch is specified. Although the recommended maximum is one second, up to four seconds is allowed. This amount of time seems long. Couple this time with the maximum allowable number of switches that can be cascaded (seven), the theoretical delay could be as much as 28 seconds! This is an eternity for an industrial control system. Fortunately, modern switches operate much faster than the standard requires.

FCS checking

A switch is required to perform a frame check sequence test on incoming frames and discard defective ones. To do this, the switch must receive the complete frame before forwarding. This means that the standard does not allow cut-through or modified cut-through operation.

Bridge addressing

The standard requires that not only must the bridge have a MAC address, each port must have a MAC address. This is unnecessary for normal switch operation. Many commercial switches do not support this requirement.

SUMMARY

Switches are classified as bridges and operate at the data link layer. They can create a much larger network diameter by segmenting the network into separate collision domains. Switches can learn from their environment and then restrict traffic only to necessary ports. This frees up other ports to initiate their own independent transmissions thereby increasing the performance over a shared Ethernet network. Repeating hubs have their place but depending upon the application, switches could provide a better solution.

REFERENCES

The Switch Book, Rich Seifert, 2000, Wiley Computer Publishing

Ethernet The Definitive Guide, Charles E. Spurgeon, 2000, O’Reilly & Association, Inc.

International Standard ISO/IEC 8802-3 ANSI/IEEE Std 802.3, 1996, The Institute of Electrical and Electronic Engineers, Inc.

ANSI/IEEE Std 802.1D, 1998, Information technology—Telecommunications and information exchange between systems—Local and metropolitan area networks—Common specifications—Part 3: Media Access Control (MAC) Bridges