Volume 7 • Issue 3
MAY-JUNE 2006

the EXTENSION

A Technical Supplement to Control Network

# Survey of Ethernet Redundancy Methods

*By George Thomas, Contemporary Controls*

*This article was edited from a paper originally presented at the ICOA 2006 conference in Shanghai, PRC.*

## Introduction

To protect against a network failure while using Industrial Ethernet, users are seeking cabling topologies that remain functional under a single cable loss. There are four popular redundancy schemes for Ethernet: Link Aggregation (Trunking), Proprietary Ring, Spanning Tree Protocol (STP), and Rapid Spanning Tree Protocol (RSTP). Each of these approaches has a set of benefits and tradeoffs, however, the industry seems to focus only on one aspect of redundancy and that is performance. How quickly does the network repair itself? This recovery time depends upon the Industrial protocols being used. Protocols such as Modbus/TCP and EtherNet/IP rely upon the TCP/IP suite of transport layer protocols, and they play a major role in how a network recovers for a single cable fault. A test was conducted using the various redundancy schemes in order to determine typical recovery times and the results were tabulated.

## Topology Options

When we discuss Ethernet redundancy schemes, we are talking about cable redundancy. How can an Ethernet network continue to function after a single cable failure? This implies there is an alternate route to carry network traffic when a failure occurs to the primary path. However, Ethernet's star topology is not conducive to cable redundancy.

## Star, Distributed Star, Tree Topology

With star topology, stations are interconnected using a wiring hub. By connecting a wiring hub to another wiring hub, we introduce the distributed star or tree topology. Wiring hubs are available as either repeating hubs or switching hubs. With star topology, end stations connect to individual ports on a wiring hub. This type of wiring is convenient to implement in a plant but still does not provide for wiring redundancy. However, the development of switched Ethernet technology can provide a solution to the cable redundancy problem.

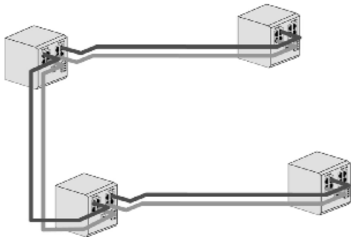
## Switched Ethernet Technology

Switching hubs offer several advantages over repeating hubs. Unlike repeating hubs that function at the physical layer of the ISO Reference Model, switching hubs operate at the data link layer. The IEEE calls this class of equipment "bridges", but they are commonly referred to as switches.

Unlike repeating hubs, switches store-and-forward complete Ethernet frames. Switches segment the Ethernet network into separate collision domains. This allows for virtually unlimited geographic expansion of a network by the simple cascading of additional switches. Switches improve network throughput by limiting traffic to only those segments that are party to the communication. Switches learn the location of end stations by observing the source address within Ethernet frames that traverse the switch. The switch takes note of the station address-port number relationship in its address table. IEEE would call the address table the "filtering database" and the process of creating assignments would be called "learning". Subsequent transmissions to a learned station would only be directed to its assigned port. This process is called "forwarding". If the switch does not know the exact location of a station, it would forward the frame to all switch ports. This is called "flooding". Since it is possible to physically move cables from one port to another, the address table could become invalid preventing a station from receiving its messages. To correct for this, address table assignments are periodically cleared forcing the relearning of the station port–assignments. This process is called "aging".

## Trunking or Link Aggregation

This redundancy approach maintains the star topology but instead of having one path between switches, two or more parallel paths are used. These multiple paths are called a trunk group and function as one larger channel. Complete frames are alternately sent down each of the parallel paths and recombined at the other end. By using multiple paths, the throughput increases with the number of separate paths. In some instances, a failed path will result in the data being diverted over the functioning paths, providing cable redundancy. This method is called "trunking". The IEEE has standardized this approach as Link Aggregation, but not all switches support this feature.



The advantage of trunking is that it provides an incremental increase in throughput as parallel paths are added. Trunk groups are not restricted to just two paths and more can be added to increase throughput. It is very easy to understand and to configure switches for trunking. Recovery time from a cable fault is extremely fast as the switches divert traffic to functioning paths. However, there will be a reduction in throughput until the cable fault condition is corrected.

The disadvantage of using trunking is that it requires the installation of additional cable. Depending upon the size of the trunk group, cable requirements can double or triple. The port count on switches increases as well which could force the purchase of larger switches. Trunking schemes are not always standardized among vendors which may restrict the purchase of all equipment from the same vendor. Although trunking supports star topology, implementing star topology becomes more involved because of the increased cable usage.

## Ring Topology and the Dreaded Loop Condition

It would appear that a ring topology would be the logical choice for redundancy since a break at any point along the ring would still leave all stations connected. However, Ethernet does not inherently support this type of topology since it would result in an endless loop condition.

Switches store-and-forward frames. A frame received on one port is forwarded to the port indicated in its address table. However, a broadcast frame that is intended to all other stations is flooded to all ports on the switch. The next switch in line will do the same and eventually the broadcast will return to the originating switch repeating the process. This situation will continue endlessly, completely consuming bandwidth, until one of the ring segments is broken. This situation needs to be avoided if ring topology is to be employed.

By using switched Ethernet technology, it is possible to wire a network in either a ring or mesh topology while guarding against the loop condition. This is accomplished by blocking those ports that will create a loop condition, and activating those same ports when a primary path failure occurs. We will study three approaches: Proprietary Ring, Spanning Tree Protocol, and Rapid Spanning Tree Protocol.

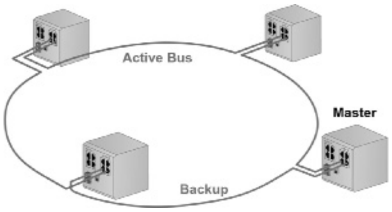
## Proprietary Ring

Several vendors have developed their own implementation of a ring topology that continues to function when one segment within the ring is broken. Typically, all switches are connected to one another through designated ring ports. One of the switches is assigned as master and blocks the port that would result in a loop condition. Each switch monitors the condition of its ring ports and reports to the master if one of its ring ports has failed. Upon receiving a report of a port failure, the master enables its backup port and instructs all switches to clear their address table. This rapid aging of each address table is to force the switches to relearn the new station-port assignments since there has been a topology change. Once the fault condition is corrected, the master is informed accordingly.

The master then informs all switches to again clear their address table; the master disables the backup port and instructs the switches to enable their primary ring ports.

The advantage of Proprietary Ring is that it is simple to understand and configure. Recovery times are very fast. Any link segment such as twisted-pair or fiber optics will work. Some implementations will even support coaxial cable. Recovery times are repeatable.

The disadvantage is that the schemes are proprietary requiring that all equipment come from the same vendor. Some implementations require a separate redundancy manager. This scheme only operates as a ring and sometimes ring topologies are not convenient to implement in a factory.



## Spanning Tree and Rapid Spanning Tree Protocols

The standardized method of supporting alternate cabling paths is explained in IEEE 802.1D Media Access Control (MAC) Bridges. Two protocols have resulted from this standard. The original was the Spanning Tree Protocol and the second is the Rapid Spanning Tree Protocol. STP is based upon timers and will operate with link segments as well as coaxial cable segments. RSTP provides much faster recovery times because its protocol examines the link status of ports and it has reduced the number of states within the spanning tree algorithm to just three: “forwarding”, “learning” and “discarding”. It will only operate with link segments. Both protocols are very complex, but they will operate with ring or mesh topologies.

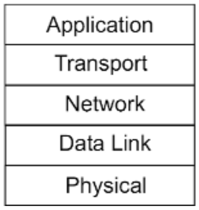
The two protocols will interoperate. Bridges communicate to one another by sending out Bridge Protocol Data Units (BPDUs). The revision number of the BPDU identifies either STP or RSTP. Within the BPDU, there is information about the topology of the network. Bridges determine the best path to the “root bridge” based upon “path cost” which is reported by the various bridges. One bridge is elected the “root bridge” (based upon the lowest value Bridge ID) while all other bridges revert to “designated bridge” status. Alternate paths are identified with one of the connected ports reverting into the “discarding” state so as not to create a loop. Upon a topology change due to a failed connection, the discarding port reverts to forwarding and the alternate path is enabled.

The advantages of STP and RSTP are that they are part of an IEEE standard that is well supported by the various vendors allowing for the mixing of vendor equipment in the field. Both protocols are not limited to just ring topology. A mesh topology will work as well. The protocols could also be used in a star topology to guard against cabling mistakes that can result in a loop. RSTP has a much faster recovery time then STP since it is not solely based upon timers.

The disadvantages to either STP or RSTP is that both protocols are very complex to understand and difficult to configure. Parameters may need to be “tuned” in the field in order to yield acceptable recovery times. STP has a very slow recovery time and represents the “old” standard. Depending upon the application, the STP recovery time may be unacceptable.

## ISO Reference Model Used in Industrial Automation

We are familiar with the seven-layer ISO Open Systems Interconnection Reference Model used to define communication tasks. The Internet version collapses this model down to five levels, and that is what best describes Industrial Automation communication. For Ethernet networks we would typically have 100BASE-TX at the physical layer, 802.3 Ethernet at the data link layer, the Internet Protocol (IP) at the network layer, either the Transmission Control Protocol (TCP) or the User Datagram Protocol (UDP) at the transport layer and finally one of the several industrial automation protocols at the application layer. Application layer examples would include Modbus/TCP, EtherNet/IP, and BACnet/IP. They differ in how they use the services at the lower layers. Modbus uses the services of TCP while EtherNet/IP uses UDP for implicit messaging. BACnet uses UDP. The type of service used could have an impact on recovery time.



TCP and UDP are both transport layer protocols that operate quite differently. TCP is connection-based guaranteeing the delivery of a message while UDP is connectionless and only provides best effort delivery services. With TCP corrupted or failed packets are automatically resent. With UDP there is no automatic acknowledgement of a successful receipt of a packet. That task is left to the application layer. Industrial automation protocols typically take this approach since it improves the real-time performance of the network. This was a consideration as we tested the various redundancy schemes for recovery responsiveness.

## Obtaining Empirical Recovery Time Data

In order to test for recovery times, we needed to create a representative network. Instead of using commercial programmable logic controllers (PLCs) and Ethernet-based input/output (I/O) devices, we used two PCs running custom programs. By using PCs, we can construct the length and type of Ethernet frame we want, and send it out as either a TCP segment or a UDP datagram. We could vary the transmission rate to simulate the performance of a PLC. The first PC is called the master. It will function as the host PLC by initiating a very simple master/slave protocol to the second PC that is functioning as a slave. It only responds to the command that came from the master. The master/slave protocol is the most common protocol used in industrial networks. Instead of the master commanding the slave to report its input status or to set the slave's outputs, we have instructed the slave to immediately repeat the command it has received. The slave is therefore functioning as an echo server; simply repeating what it has heard as fast as it can. The master awaits the response and matches it against the command to verify the message integrity. Once confirmed, another command is immediately sent. By fully understanding the average time for a response, it becomes very easy to determine the time for an abnormal response. This will be the response after removing one of the forwarding links.

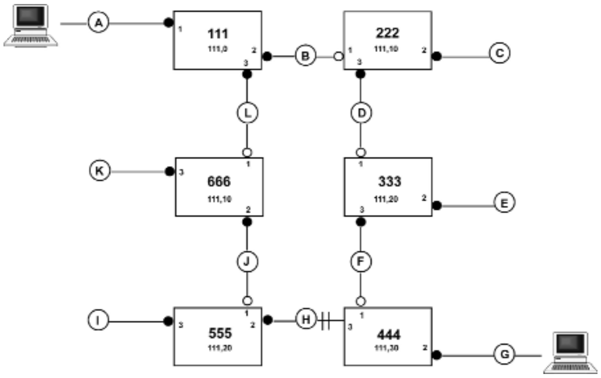
Although the above plan works well for using UDPs, we needed to change the application program in order to handle TCPs. For the TCP portion of our test, we completely relied upon TCP to acknowledge the successful transmission of the messages. Only after a successful receipt of the command message, would the echo be sent. We observed later that this change impacted the recovery times.

### *Ring Topology Instead of Mesh*

The next decision was to select the network topology and the number of switches in the network. We elected to only use managed switches that were capable of being configured for STP, RSTP, trunking or proprietary ring. We decided to use the same ring topology that was shown in the 802.1D-2004 standard for RSTP. Although STP and RSTP can function in either a mesh or ring topology, we did not want to change the topology when we tested for proprietary ring. In this way we would have consistent data. Besides, the ring topology tends to be the more popular selection when incorporating redundancy schemes. This presented a slight problem with the trunking test since loops are not allowed. For trunking, we broke the ring by removing the same segment that was being used as the alternate segment. In this way the same number of switch hops (in our case four) would occur during communication as would be experienced with STP, RSTP and proprietary ring. It must be remembered, however, that with trunking a segment in the diagram really means two separate paths. So when we test the performance after a break, we mean that we removed one of these paths when testing the trunking scheme. Like the diagram, we used six managed switches. We could have used more but we felt this would be a representatively sized network. We set all ports to 100 Mbps full-duplex with PAUSE enabled. We chose copper cabling for convenience although we could have used fiber optics as well.

### *Normal and Alternate Paths*

This network diagram was taken from the RSTP section in 802.1D. In this diagram there are six switches each shown with a unique Bridge ID. Since Bridge 111 has the lowest numerical bridge value, it becomes the root bridge in STP and RSTP testing. When conducting either the trunking or proprietary ring tests, the root bridge has no special significance. There are three ports on each switch numbered from 1 to 3.



A port with a solid black dot indicates that the port will be "forwarding frames" (normal operation) away from the root bridge. A port with an unfilled circle means that the port is "forwarding frames" towards the root ridge. Attached to segment A will be our master PC and attached to segment G is our slave PC. You will notice that port 3 on Bridge 444 is "discarding". What this means is that it is receiving traffic, but it does not pass the traffic on to its switch fabric because to do so would create a loop. However, this port is attached to a healthy segment H that could form the alternate path in the event of a segment failure somewhere else around the ring. Therefore under normal operation, a transmission from the master PC will travel through switches 111, 222, 333 and then 444 before it gets to the slave PC. The slave to master response would travel through the same path in reverse order. If a break occurs at segment D, the network will reconfigure such that a transmission from the master PC will now travel from 111, 666, 555 and then 444 before it gets to the slave PC.

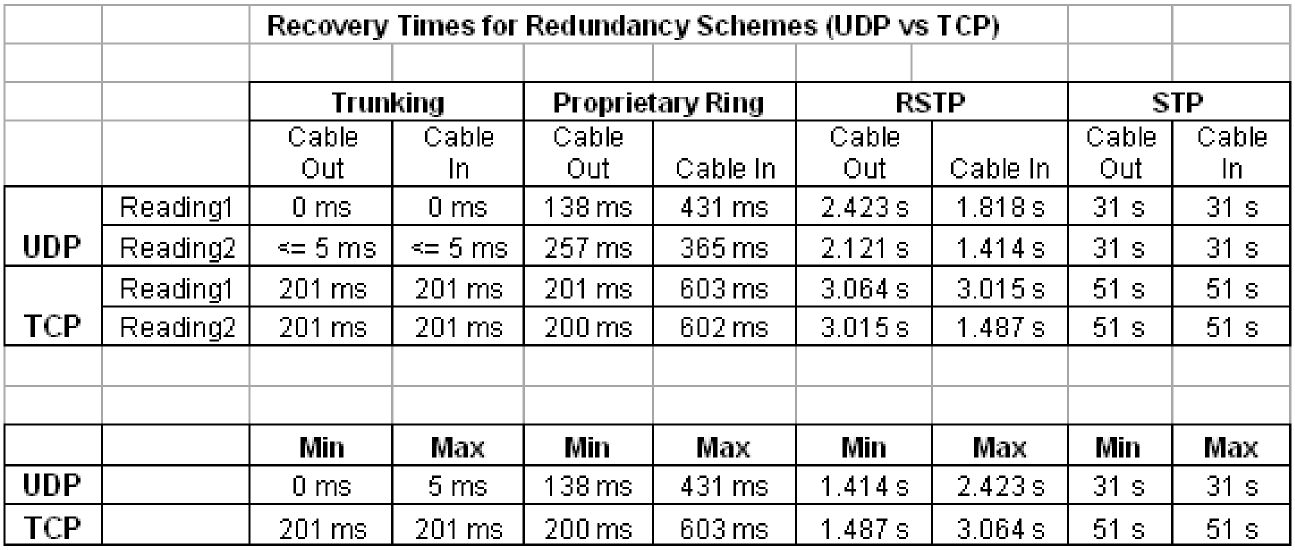
## Test Results

Once the representative network was operating, we simulated a break by the removal and insertion of a cable. Both instances resulted in a topology change so we measured the time for the network to recover and resume normal operation. We first ran the UDP test followed by the TCP test for all four redundancy schemes. In each case, we attempted two successive readings. The results are in the table below.

Trunking provided the best results with an amazing 5 ms recovery time using UDP. We had to make several readings just to capture the failure since we could not send messages as fast as the recovery time. Proprietary ring came in second with a 138–431 ms recovery time again using UDP. RSTP was not as rapid but still very fast with recovery times in the range of 1.4 to 2.4 seconds. TCP results were worse. STP was a distant fourth with a 31 second recovery using UDP and 51 seconds under TCP! Clearly, TCP had an impact on recovery time. However, to be fair to TCP, recovery times could have been improved. Once an application program realizes that it lost communication, it could have broken the TCP connection and re-established the connection in order to avoid the lengthy time-out process.

## Summary

Even though STP provided the poorest performance, it does not mean it is unusable. It depends upon the application. Applications such as building automation and some process control applications may continue to function even if communication is lost for 30 seconds or more. Other applications, especially in safety or security, may not tolerate any disruption thereby causing controller shutdowns. There may be no practical solution if several Ethernet frames are lost during the recovery process. Users should know the requirements of the application before insisting on a particular redundancy scheme.



| | | Recovery Times for Redundancy Schemes (UDP vs TCP) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | |
| | | Trunking | | Proprietary Ring | | RSTP | | STP | |
| | | Cable Out | Cable In | Cable Out | Cable In | Cable Out | Cable In | Cable Out | Cable In |
| UDP | Reading1 | 0 ms | 0 ms | 138 ms | 431 ms | 2.423 s | 1.818 s | 31 s | 31 s |
| | Reading2 | <= 5 ms | <= 5 ms | 257 ms | 365 ms | 2.121 s | 1.414 s | 31 s | 31 s |
| TCP | Reading1 | 201 ms | 201 ms | 201 ms | 603 ms | 3.064 s | 3.015 s | 51 s | 51 s |
| | Reading2 | 201 ms | 201 ms | 200 ms | 602 ms | 3.015 s | 1.487 s | 51 s | 51 s |
| | | | | | | | | | |
| | | | | | | | | | |
| | | Min | Max | Min | Max | Min | Max | Min | Max |
| UDP | | 0 ms | 5 ms | 138 ms | 431 ms | 1.414 s | 2.423 s | 31 s | 31 s |
| TCP | | 201 ms | 201 ms | 200 ms | 603 ms | 1.487 s | 3.064 s | 51 s | 51 s |

www.ccontrols.com

Past issues of the copyrighted Extension are available. Please visit our web site www.ccontrols.com. Select Support and click on Extension Archive.

## References

**IEEE Std 802.1D™-2004**
IEEE Standard for Local and Metropolitan Area Networks: Media Access Control (MAC) Bridges