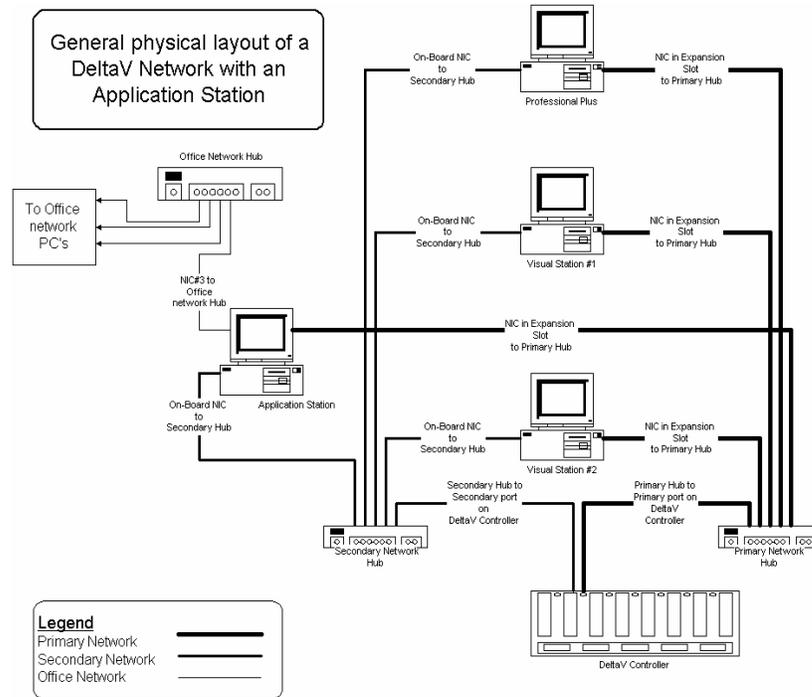




Ethernet Fault Tolerance and Redundancy

This document discusses implementing a fault-tolerant Ethernet network and the DeltaV approach.



© Emerson Process Management 1996—2007 All rights reserved.

DeltaV, the DeltaV design, SureService, the SureService design, SureNet, the SureNet design, and PlantWeb are marks of one of the Emerson Process Management group of companies. All other marks are property of their respective owners. The contents of this publication are presented for informational purposes only, and while every effort has been made to ensure their accuracy, they are not to be construed as warranties or guarantees, express or implied, regarding the products or services described herein or their use or applicability. All sales are governed by our terms and conditions, which are available on request. We reserve the right to modify or improve the design or specification of such products at any time without notice.





Contents

Introduction	4
Fault Tolerant Architectures	4
Redundancy	4
Media Redundancy	4
Network Node Redundancy	6
DeltaV Control Network	7
Communication Diagnostics	7
DeltaV Network Services	7
Summary	9
References	9



Figures

Figure 1 Ring using STP	5
Figure 2 Ring using STP with failure in one path	5



Introduction

A modern control system typically is deployed in a highly distributed manner, with communications between nodes thus becoming a critical part of the system architecture. This means that some form of fault tolerance of this communications network is required in order to achieve satisfactory system availability. High availability, achieved through redundancy and fault tolerance, is a critical component of many industrial installations. While the loss of an enterprise network for a few minutes is inconvenient, losing an industrial network can have disastrous consequences. By using a standards-based solution that supports multi-vendor implementations, users enjoy highly reliable systems, reduced costs of deployment, and a guaranteed upgrade strategy as needs evolve. This whitepaper discusses several different fault tolerant architectures for the Ethernet communications network, as well as the approach taken by the DeltaV system.

Fault Tolerant Architectures

In order to understand the effectiveness of differing fault tolerant architectures, we also need to consider whether the desired criterion is to continue operation or to fail to a safe state. Different architectures are more suited to continued operation, while others are more suited to fail-safe. Combinations are also possible and are sometimes employed in an attempt to get the best of both worlds, typically at additional cost. It is also important to consider the cost impact of the architecture, together with the ease of implementation. A more complex—and thus more expensive approach—may not be warranted, and it may actually negatively impact the availability. Additionally a proprietary networking solution may have long-term support consequences. (Is the network manufacturer able to provide the same long-term support guarantees that the system vendor provides?)

Redundancy

Redundancy is typically achieved by duplicating components, although in some circumstances a diverse redundancy scheme is employed. A non-diverse redundancy scheme works well in protecting against external faults if the fault does not have a common-cause effect on the redundant components. For example redundant media are typically employed for data highway communications, and this is very effective in preventing problems due to accidental damage of the cable—BUT only if the cables are routed through different paths. The often-discussed “forklift through the cable tray accident” will not be prevented if the cables are in the same cable tray. The DeltaV system deploys a dual-star topology to create an automation system network that is readily available and also reliable, but the cost requires duplication of the network equipment such as switches, etc.

Media Redundancy

Media redundancy, which involves forming a backup path when part of the network becomes unavailable, is another approach that is often chosen for automation. One of the technologies developed for media redundancy is called IEEE 802.1D Spanning Tree Protocol, or STP for short. STP has proven in general use over many years to be interoperable, and commercial systems using products from multiple vendors are routinely implemented. Standard STP supports redundant configurations of any type: meshes or rings or combinations. Many industrial implementations use an Ethernet ring topology with backup paths. Prior to the development of STP, it was not possible to create an Ethernet ring topology since loops in an Ethernet network are not allowed. While STP has demonstrated its capability to provide loop-free connectivity and fault-recovery through the determination of an alternative path, its convergence time is 30 to 50 seconds—much too long for many industrial applications. What STP does is to identify one of the switches in the network as the “root switch” of the network, and then automatically

block packets from traveling through any of the network's redundant loops. In the event that one of the paths in the network is disconnected from the rest of the network, the STP automatically readjusts the topology and uses the redundant path.

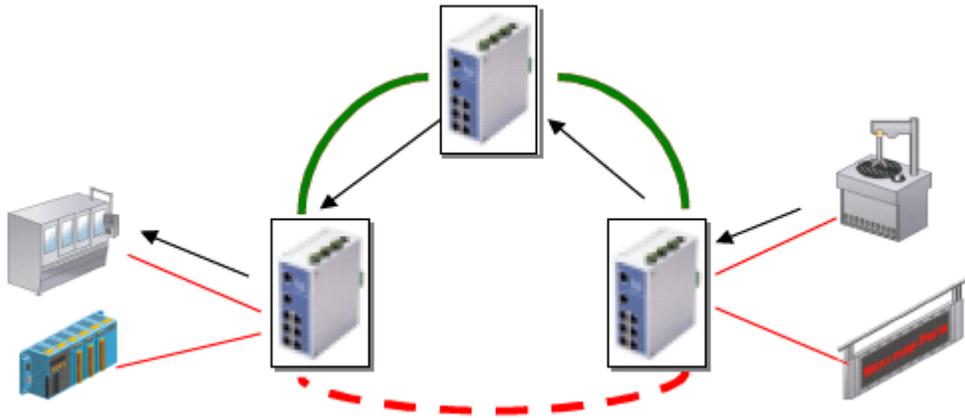


Figure 1 Ring using STP

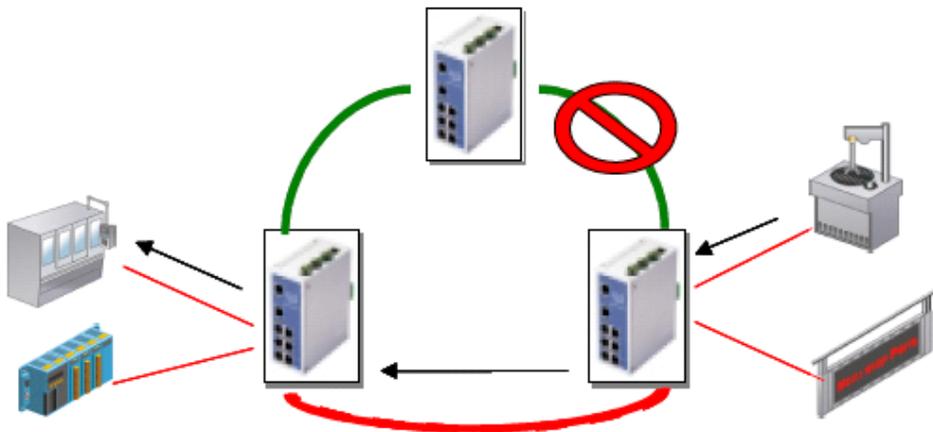


Figure 2 Ring using STP with failure in one path

Although IEEE 802.1D STP has solved some limits of Ethernet network technology, it also has limitations, including the lower convergence speed mentioned previously. The seeds of STP's weaknesses reside in its strengths: It is not inherently ring-oriented, and the complexity that allows it to support a variety of topologies limits its performance in a relatively simple redundant ring, as would typically be deployed in an industrial control system.

When there is a fault in a ring, the obvious solution is to treat the interrupted ring as two separate strings until the fault is repaired. The time it takes standard STP to make an analysis of the situation that conforms to the standard is both unnecessary and a weakness when fast fault recovery is an objective. For this reason, IEEE 802.1W Rapid



Spanning Tree Protocol (RSTP) was developed. This newer protocol has all the advantages of IEEE 802.1D, but in addition provides higher performance. RSTP can also work with legacy STP protocols, and start a migration delay timer of 3 seconds. RSTP can normally respond in one to three seconds in small- or medium-sized meshes and some small rings. The benefit RSTP brings is that it pre-calculates routes to fall back upon when a port, switch, or link failure is detected. Because it can distinguish between point-to-point and shared links, it can reduce the overhead on point-to-point and “edge” links where the number of rerouting options is predetermined. The technologies mentioned above made media redundancy with high performance not only possible, but also feasible. RSTP does have a theoretical limit of 31 switches in a ring topology, which may not be sufficient for some industrial ring applications. For this reason, many Ethernet device manufactures are developing proprietary protocols based on 802.1W to meet the fast recovery time required and to allow for the large numbers of switches often needed in industrial automation.

The benefit of going proprietary is more rapid fault resolution than waiting for a standards-based solution. The downside is the risk associated with becoming locked in to a single source. Ethernet switches learn MAC addresses in order to switch packets to their destination port and save the addresses in their memory as long as they are active. If a MAC address ceases to be active, it is aged out of the switch memory after a few minutes. This switch-address-aging delay presents a problem when a LAN needs to be reconfigured quickly. While repeaters (hubs) have no address buffer and, therefore, do not create a recovery bottleneck, the switch’s stored addresses prevent packets from going via a new recovery route until the addresses have aged out of switch memory.

There is no benefit to having a fast ring recovery technique if the switch members of the ring prevent Ethernet traffic from moving to the recovery traffic path. Different switch vendors implement different address buffer aging times. In a multi-vendor implementation, the slowest aging time in the recovery path will govern the ring recovery time. It is critical to build in switch address buffer aging times when calculating the time-to-recovery in a redundant ring.

For proprietary ring solutions, there is limited—or no—interoperability with other products on the market. To solve this problem, several vendors have contributed to a new IEC committee in hopes of driving a standard. This means that once there is a standard, the proprietary solutions will likely need to change their product to meet the standard. This results in the solution being more costly, both in initial purchase price and in the lifecycle costs as standards-based solutions evolve. Will there be an easy way to deploy upgrade path? This is the question that end users need to consider.

Network Node Redundancy

After successfully implementing media redundancy, another problem is how to include other likely failures in the entire control system. Is it more likely for a passive wire or fiber to fail or for an active piece of electronics (switch or media converter)? For this reason, switches that are connected to critical devices such as controllers need to set up dual network paths, i.e. one needs to deploy a second Ethernet switch. To keep the system running when a network problem occurs, a controller that supports two Ethernet interfaces to connect both redundant switches is required. Once the decision has been made to use redundant network switches, the needs for a ring topology for fault tolerance become significantly diminished.

A completely redundant system consists of redundant switches, redundant communication ports, and redundant device pairs. All Ethernet devices and workstations are connected to both independent network architectures. Complete system redundancy can form an extremely reliable network that minimizes data loss and has fast recovery time.



DeltaV Control Network

The DeltaV control network (ACN) is based on standard Ethernet with optional redundant media. This provides a high degree of fault tolerance due to the inherent retries in TCP/IP. We have also employed additional message checking in the application layer. All devices on the control network have dedicated Ethernet connections including redundant controllers. The network is deployed using standard hubs/switches. This provides an additional layer of security, whereby a fault on one device (or cable) will be isolated by the hub/switch, thus preventing the fault's propagating to other nodes. ACN redundancy is accomplished by using two separate Ethernet communications links or networks. The primary communication link is the preferred communications path. The secondary link is used only for DeltaV traffic if the primary has failed (the secondary is configured to carry DCOM and non-DeltaV traffic when both are available. If the secondary fails, this traffic is routed to the primary). Communications switchovers are performed on a per-node basis. For example, if Node A is communicating with Nodes B and C and the primary link to Node C fails, Node A will continue to communicate with Node B on the primary but will switch to the secondary link to communicate with Node C. Any time a communications switchover occurs, an event is generated to notify the operators that a communications link switchover has occurred.

A redundant DeltaV network is configured to avoid problems if an Ethernet network goes down. There are no shared wires, no packet routing, and no traffic that is common between these two networks. During normal operation, the primary network carries the traffic between the DeltaV nodes. Windows NT/XP communication (i.e. file transfers, printing, and any other non-DeltaV traffic) will default to the secondary network, thus leaving the primary network open for DeltaV communication only. If the secondary network becomes unavailable, the primary will be used for all communication. The amount of traffic from Windows NT/XP communication should be low and infrequent, but is necessary. More information on the control network and suitability of Ethernet for control can be found in another whitepaper.

Note: Within the DeltaV ACN LAN, only the certified vendor and model of networking equipment (hubs and switches) may be used. We do not support or permit the use of non-certified network hardware (including the NIC cards in the workstations) within the DeltaV ACN. The use of routers within the DeltaV ACN is expressly not permitted.

Communication Diagnostics

All DeltaV PC workstations and controllers provide detailed diagnostic information about the status of the communications subsystem in that device. Each node will support detailed integrity information about the status of the ACN communications links, node connections, Ethernet statistics, and communications stack diagnostic information. A diagnostic "ping" (check for communications) is supported as part of each node's diagnostics.

The communications diagnostics provide three levels of diagnostics information.

- Network Communications Status—basic diagnostics for the whole network
- Node Communications Status—diagnostic information for a single node
- Node Connection Statistics—diagnostic information for a single node connection

DeltaV Network Services

The ACN communications system is made up of several distinct parts. The Low Level Communication (COMM) is responsible for the interface with the TCP/IP sockets, redundancy, node connections, and the actual transmission of the messages across the wire.



COMM provides the following services:

- 1) **Connection Management**—independent peer-to-peer node connections. Each pair of communicating nodes determines the ability to communicate over primary and secondary interfaces. System configuration verification on “connection establishment connection requests” must match all configurable parameters before the request is accepted.
- 2) **Message Delivery Services** –
 - a. *Guaranteed message delivery*—Messages are retransmitted until positively acknowledged or a timeout occurs.
 - b. *In sequence message delivery*—Messages are always received in the order in which they were sent. Per packet message verification is enforced. Each message must match the expected sequence number and certain data fields.
 - c. *Duplicate message detection*—Duplicate messages received out of order will not be accepted.
 - d. *Flow control*—A temporary shortage of receive buffers will notify peers to stop sending until this condition clears. Peers are notified when the buffer shortage clears.
 - e. *Timeout and retry*—After a message is sent, a positive acknowledgment must be received within the timeout period or the message is retried.
- 3) **Link Detection and Switchover** –
 - a. *Link Detection*—Link errors are detected when messages from a peer are no longer received. A switchover to the secondary interface will occur if an error is detected on the primary interface. A switchover back to the primary interface will occur when the error on the primary interface is corrected. Event notification for comm failures and switchover errors between peers are logged in the DeltaV event journal.
 - b. *Active and Standby connection management*—Active and standby connections for redundant nodes (Controllers). The comm layer handles establishing connections to active or standby nodes based on the application request.
 - c. *Reestablishment of node connections after a redundant node switches over*—Reestablishment of active node connections after a node switches from active to standby. The new active node will immediately inform all peers of the switchover.
 - d. *Link error detection and switchover for non-DeltaV applications*—Communication integrity is maintained among all DeltaV workstations using IP-based communications. Applications will continue uninterrupted as long as one good link exists (e.g. configuration applications are communicating over DCOM).
- 4) **Time Synchronization**—Time synchronization using NTP workstations and controllers are synchronized using standard Network Time Protocol.
- 5) **Communications Diagnostics**—Keeps track of communications integrity and statistical information. Responds to requests for communications diagnostic data and provides a mechanism to ping a DeltaV node on the network to verify that it exists and is capable of communicating.



Summary

A single ring of Ethernet is better than a single Ethernet network in topology, but a real redundant Ethernet network like the DeltaV system has is better than a ring. Every node on the DeltaV redundant control network plays a role when redundant switchover occurs. And comparing with the DeltaV redundant network, a double ring is meaningless and expensive in capital, operational and maintenance costs. Lifecycle costs must also be taken into consideration.

References

1. Charles Spurgeon's Ethernet website - <http://www.ots.utexas.edu8080/Ethernet/>.
2. Redundancy in automation—Moxa
3. Redundancy with Standards in Industrial Ethernet LANs GarrettCom, Inc