HIRSCHMANN
A BELDEN BRAND

White Paper

WP 1003HE

Media Redundancy Concepts

High Availability in Industrial Ethernet

Contents

1. Introduction

The general idea of media redundancy and redundant paths is almost as old as the use of Ethernet for industrial communications, and so is the dilemma that – by definition – Ethernet technology's broadcast nature does not permit physical loops and therefore effectively forbids redundant communications paths.
However, faulttolerance, which necessitates the use of redundant structures, is a vital basic requirement of very many automation systems.
This means that the use of Ethernet for automation technology applications calls for protocols that are able to resolve the physical loops generated by the introduction of redundant pathways.
To facilitate the use of redundant communications structures in office environments, the IEEE (Institute of Electrical and Electronics Engineers) specified the spanning tree protocol (STP), which was published in the 802.1D 1990 standard. For the first time this enabled all Ethernet switches to employ an algorithm to facilitate interconnected network structures, albeit with switchover times of the order of many tens of seconds. Further protocols based on the underlying STP mechanisms were subsequently developed, and these were better tailored to the specific requirements of an industrial environment, in particular with markedly reduced switchover times.
This white paper will give you an overview of the current state of the technology and its solutions and also sketch a number of specific applications.

2. Reasons for media redundancy

Media redundancy is primarily used to avoid single points of failure in industrial communications networks. Wherever there is a single point of failure it is possible for the communications network, for instance in an automated production line, to be completely disabled by a single technical fault. The consequences of such a failure can potentially be extremely costly. If redundant structures are used then a single failure merely causes the network to fallback to a degraded state. Communications via the network remain viable, and the redundant system makes it possible for a repair to be carried out to restore the previous fault-free state. You will find additional detailed information about high-availability systems, media redundancy, failure and repair models in [1].
Redundant network structures are used for two separate purposes:

1. Load balancing: The data traffic over a network path within a specific interval is greater than the bandwidth that a single data cable is able to handle. Introducing additional redundant connections increases the effective bandwidth of the original connection. The IEEE's link aggregation control protocol (LACP) [2] is typically used for this purpose.
2. Fault tolerance: Additional media connections between network subscribers are introduced to enable the system to switch over to a secondary network path in the event of a failure in the primary path.

Although the second case may be contained within the first, the use of media redundancy in an industrial environment is usually restricted to the second case. In an industrial network, fault tolerance is far and away more important than load distribution, and this is why most of the protocols used in this field specialize in ensuring high availability.

High availability is a crucial requirement for all automation systems, whether in manufacturing, processing or substation automation applications.
Component failure, which can never be entirely ruled out, needs to be dealt with in such a way as to minimize its impact on the system as a whole.

3. Basic requirements for industrial use

One fundamental requirement for any Ethernet network is the avoidance of loops. There must at all times be exactly one path between a message source and the corresponding sink. Any loops will result in data packets that circulate endlessly and eventually overload the network, which is why Ethernet does not permit alternative active paths to its devices. But media redundancy needs these alternative paths. Resolving this conflict calls for a protocol to monitor the redundancy. Such a protocol must guarantee that, at any one time, there is only a single logical path to each device, even if there are a number of physical pathways. The protocol achieves this by making sure that only one of the possible pathways is active at any one time and all the others are in standby mode.
The solution, which was realized for the first time with STP, depends on monitoring the links, detecting interruptions in communications and switching to an alternative path as soon as a failure is detected. Note that this principle means that communications will be interrupted for a certain time, because the failure first needs to be detected before the network can be switched over to the alternative path and communications are restored.
Depending on the complexity of the network, the duration of such interruptions may be difficult to predict.

The following fundamental requirements apply to media redundancy protocols in an industrial environment:

1. Switchover-time determinism: in the event of a failure, the time the protocol needs to switch from the primary logical path to a secondary alternative path and to restore communications must be predictable.

2. Installation requirements: If using the protocol and/or complying with required switchover times impose any constraints on the installation, for example the physical topology or the maximum number of useable network switches, then these must be clearly specified.

3. The protocol must be based on a standardized method. This is the only way of guaranteeing transparency, compatibility and hence security of investment.

The first requirement is absolutely essential for the use of automation technology or related applications. A media redundancy protocol can be used only where reliable and calculable figures are available to specify the absolute worst-case upper limit for network switchover time in the event of a failure. This is the only way of ensuring that the network will fulfill the requirements of the application that is using it as a transmission medium:
If the media redundancy protocol can switch over fast enough to enable the protocol traffic and application to continue operating without impairment, then its redundancy mechanism is transparent to the application functionality and the timing requirements are fulfilled.

4. Technologies and solutions

4.1 RSTP/MSTP – Rapid/Multiple Spanning Tree Protocol

4.1.1 RSTP/MSTP overview

Over the last few years the spanning tree protocol mentioned earlier has been largely superseded by the rapid spanning tree protocol, RSTP. This is an optimized version of STP that was definitively described in the IEEE 802.1D 2004 standard [3]. RSTP implementations operate in a variety of topologies, support a higher number of switches and achieve improved switchover times of the order of about one second. However, RSTP still does not guarantee deterministic failure behavior. Reaction times depend on the location in the network where the failure occurs, and also on the approach taken by the individual implementation.
For this reason there have been a number of attempts to optimize RSTP by restricting it to ring topologies and using fixed predefined parameters. To date, these optimizations have made it possible to demonstrate switchover times of the order of 100 ms or less (see 4.1.2).
The rapid spanning tree protocol, as its name implies, creates a tree structure from the connections between the Ethernet switches and disables all those paths that are not a part of the active tree.
This results in exactly one active path between any two devices. This protocol uses what are called bridge protocol data units (BPDUs) to communicate between the switches. One root bridge is defined as the root of the tree, and the optimal network paths are determined from there. If the network is changed in any way, for instance by the failure of a physical connection, this is reported to the network by means of topology change notification BPDUs. The response to this is to recalculate the tree, activate the appropriate alternative paths and thus restore communications.

MSTP [4] is a further development of RSTP and works on the same principle. However, while RSTP operates independently of virtual local area networks (VLANs), MSTP always operates within VLANs and therefore facilitates more flexible network structures, for instance in order to implement load balancing over a variety of VLANs and network paths. MSTP and RSTP are mutually compatible and can be used together in a single network structure.

4.1.2 Use in ring structures

If the topology is restricted to a ring, then it is possible to achieve deterministic and predictable switchover times with RSTP, provided the RSTP timing of the switches is known. The IEC 62439 1 standard contains a sample calculation that also demands additional protocol restrictions. For example, to prevent disruptive influences from outside the ring, the RSTP may not be configured on switchports other than ringports.
Since RSTP was not primarily developed for ring topologies its design does exhibit a number of disadvantages compared to the MRP described in section 4.2.
For network devices that support both MRP (with a parameter set of 200 ms or better) and RSTP, and have no installation requirements that prescribe specific protocols, MRP is preferable to RSTP.
It should also be noted that RSTP possesses built-in overload protection to prevent individual network segments from being overloaded by large numbers of event-driven BPDUs. In a worst-case situation this overload protection has the effect of greatly increasing the reconfiguration time caused by lost BPDUs, up to the order of seconds. This restriction is less apparent in ring structures because of the less flexible topology, but it may still occur. And it may happen quite frequently in meshed networks, particularly in the case of complex topologies with a high number of switches and media connections.

4.1.3 Use in meshed networks

One great strength of RSTP is its support for all kinds of meshed topologies. The resulting flexibility regarding the installation is a clear advantage over the stringent restrictions that are imposed by ring protocols such as MRP and ring installations.
However, this flexibility harbors one great disadvantage, namely the reconfiguration time, which for an interconnected network will depend – among other things – on the complexity of the network topology and the location in the network at which the failure occurred. Since RSTP, unlike MRP, is a decentralized protocol, it may also provoke highly unpredictable race conditions in the establishment of new communications paths, particularly when choosing a new root bridge. This gives rise to network reconfiguration times that can be estimated only very roughly, and this does restrict the use of RSTP, particularly in meshed networks.
In the case of meshed networks with very little complexity (such as ring networks with two or three additional loops or subrings), a detailed analysis can make it possible to determine upper limits, but these will always need to be worked out individually. Unlike with the protocols MRP, HSR and PRP, it is not possible to make a general statement.
One method of determining reconfiguration times on the basis of specific application scenarios was worked out by Hirschmann/Belden in the course of the international standardization process for the next revision of the IEC 62439 1 standard.

4.2 MRP – Media Redundancy Protocol

One protocol that particularly addresses industrial applications is the media redundancy protocol, MRP. This protocol is described in the IEC 62439 2 standard, which is the industry standard for high-availability Ethernet networks. MRP is specified only for ring networks with up to 50 devices, and guarantees fully deterministic switchover behavior. Its absolute worst-case upper limit for switchover times in response to a failure are 500 ms, 200 ms, 30 ms or as low as 10 ms, depending on the chosen parameter set.
Typical switchover times for MRP vary between half and a quarter of these worst-case times. Thus, under typical network load conditions, an MRP ring that is configured for a 200 ms worst case will need between 50 ms and 60 ms to switch over from the primary to the secondary path; under typical conditions an MRP ring with a 10 ms switchover time will react correspondingly faster.
Every MRP node requires a switch with two ringports connected to the ring.
Under MRP, one of these nodes functions as a media redundancy manager (MRM).
The MRM monitors and controls the ring topology so that it can react to network failures. It does this by sending Ethernet redundancy test frames to one ringport and receiving them at the other, and vice versa.
In a non-failure state, the MRM blocks all network traffic on one of its ringports, with the exception of MRP protocol traffic.
At a logical level, this converts the physical ring structure to a linear structure for ordinary network traffic, thus avoiding loops.
If the MRM fails to receive its test frames, indicating a transmission failure in the ring – for example because of a device failure or a defective media connection – then it will open the previously blocked stand-by ringport to normal protocol traffic.
All the devices will then be accessible via the secondary network path.

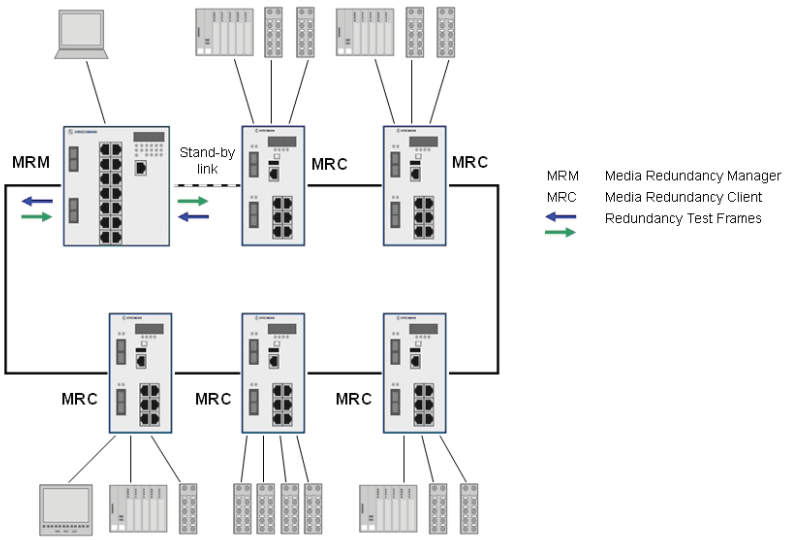


Figure 1-MRP ring

All the other nodes in the ring have the role of media redundancy clients (MRC).
An MRC conveys the redundancy test frames fed into the ring by the MRM from one ringport to the next.
It also reacts to any received reconfiguration frames (topology change) from the MRM, detects changes in the state of its port and reports this to the MRM. If such a state change report reaches the MRM before it has been able to detect the ring failure on the basis of missing test frames, then it uses the information received from the MRC to detect the failure. This ensures that the switchover in the MRM from primary to secondary network paths is always carried out within the shortest possible time.
This flexibility in the matter of switchover times and the distinction between the dedicated manager (MRM) and the resource-efficient clients (MRCs) enable the MRP ring to cover a very large number of practical requirements and be optimally configurable to suit them.

4.3 PRP – Parallel Redundancy Protocol

Although a fast MRP ring can now cover a very large number of requirements, there are still applications that cannot tolerate any switchover time at all.
To fulfill such requirements we need to take an entirely new approach to the question of guaranteed high availability.
The basis of this new approach to network redundancy is to have two independent active paths between two devices. The sender uses two independent network interfaces that transmit the same data simultaneously. The redundancy monitoring protocol then makes sure that the recipient uses only the first data packet and discards the second. If only one packet is received, the recipient knows that a failure has occurred on the other path.
This principle is employed by the parallel redundancy protocol (PRP), which is described in the IEC 62439-3 standard. PRP uses two independent networks with any topology and is not limited to ring networks.
The two independent parallel networks may be MRP rings, RSTP networks and even networks without any redundancy at all. The principal advantage of PRP is its interruption-free switchovers, which take no time at all to switch over in failure situations and thus offer the highest possible availability. Naturally this applies only provided both networks do not fail simultaneously.
PRP is implemented in the end devices, while the switches in the networks are standard switches with no knowledge of PRP. An end-device with PRP functionality is called a double attached node for PRP (DAN P) and has a connection to each of the two independent networks. These two networks may have the identical structure or may differ in their topology and/or performance.
A standard device with a single network interface (single attached node, SAN) can be connected directly to one of the two networks. Naturally, in this case, the device will have no redundant path available in the event of a failure. A SAN can alternatively be connected to a redundancy box (RedBox) that connects one or more SANs to both networks. SANs do not need to know anything about PRP, they can be standard devices.
In many applications only critical equipment will need a dual network interface and less vital devices can be connected as SANs, with or without a redundancy box.

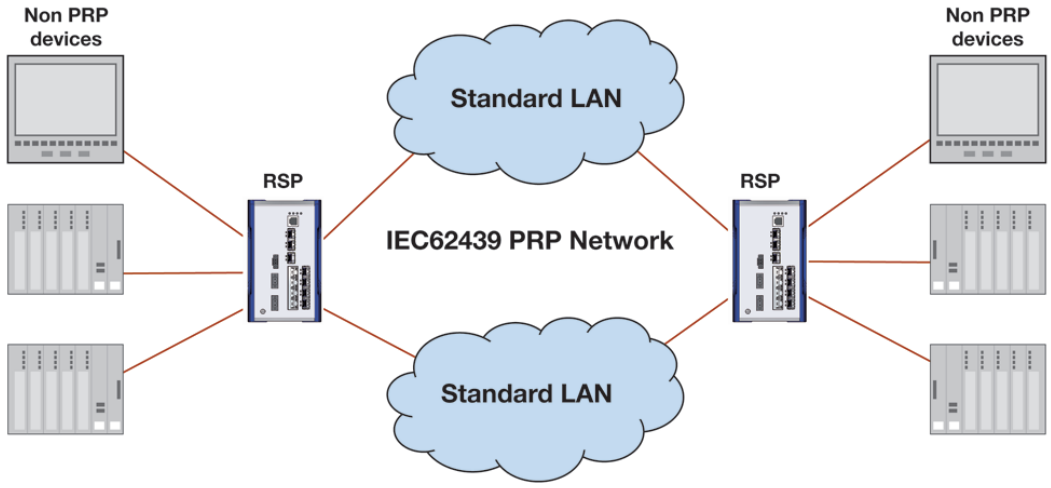


Figure 2 – Identical data packets are transmitted simultaneously to both networks

A DAN P implementation controls the redundancy and deals with duplicates. When the upper layers receive a packet for transmission, the PRP unit sends this frame to the network via both ports simultaneously. When these two frames traverse the two independent networks they will normally be subject to different delays on their way to the recipient. At their destination the PRP unit passes the first packet to arrive to the upper layers, i.e. to the application, and discards the second one. The interface to the application is thus identical to any other Ethernet network interface.

The redundancy box implements the PRP protocol for all the attached SANs and thus operates as a kind of redundancy proxy for all types of standard equipment.

Duplicates are recognized by means of the redundancy control trailers (RCT) introduced into each frame by a PRP connection or RedBox. In addition to a network identifier (LAN A or B) and the length of the user data contained in the frame, these 32-bit identification fields also contain a sequence number that is incremented for each frame sent by a node. A RedBox or DAN P connection can thus recognize duplicates, and if necessary discard them, on the basis of the clearly identifiable features contained in each frame (physical MAC source address and sequence number).

Since the RCT is inserted at the end of the frame (see Figure 3), all the protocol traffic can still be read by SANs, which interpret the RCT merely as additional padding with no significance. This means that a SAN that is connected to a PRP network directly, i.e. without a RedBox, is able to communicate with all the DAN Ps and with any SANs in the same network (either A or B). It lacks only connections to the nodes of the other network, because a DAN P does not pass any frames from one LAN to the other one. PRP switchover times fulfill the very highest demands, and it is also extremely flexible as regards network structure and possible topologies, but it does always need twice the installed infrastructure of switches and other network components.

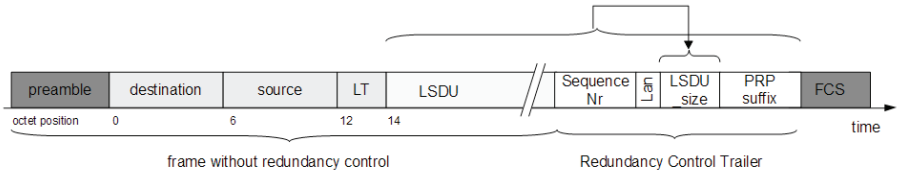


Figure 3 – PRP frame format with no VLAN tag (extract from IEC 62439 3)

4.4 HSR – High Availability Seamless Redundancy

High availability seamless redundancy (HSR) is a further development of the PRP approach, although HSR functions primarily as a protocol for creating media redundancy while PRP, as described in the previous section, creates network redundancy. PRP and HSR are both described in the IEC 62439 3 standard.
Unlike PRP, HSR is primarily designed for use in (redundantly coupled) ring topologies. Like PRP, it uses two network ports, but unlike PRP a HSR connection incorporates a DAN H (double attached node for HSR) that connects the two interfaces to form a ring (see Figure 4).

A frame from the application is given an HSR tag by the HSR connection.

Like the PRP RCT, this contains the length of the user data, the port that transmitted it and the sequence number of the frame.

However, unlike PRP, the HSR header is used to encapsulate the Ethernet frame (see Figure 5). This has the advantage that duplicates of all frames are recognized in all devices as soon as the HSR header has been received. There is no need to wait for the whole frame and its RCT to be received before a duplicate can be recognized as such. This means that, similarly to cutthrough switching, individual HSR connections and RedBoxes can begin forwarding the frame to the second ringport as soon as its HSR header has been completely received and duplicate recognition carried out. Each HSR node takes from the network all frames that are addressed only to it and forwards them to the application. Multicast and broadcast messages are forwarded by every node in the ring and are also passed to the application.
In order to prevent multicast and broadcast frames from circulating for ever, the node that initially placed the multicast or broadcast frame on the ring will remove it as soon as it has completed one cycle (see HSR data flow in Figure 4).

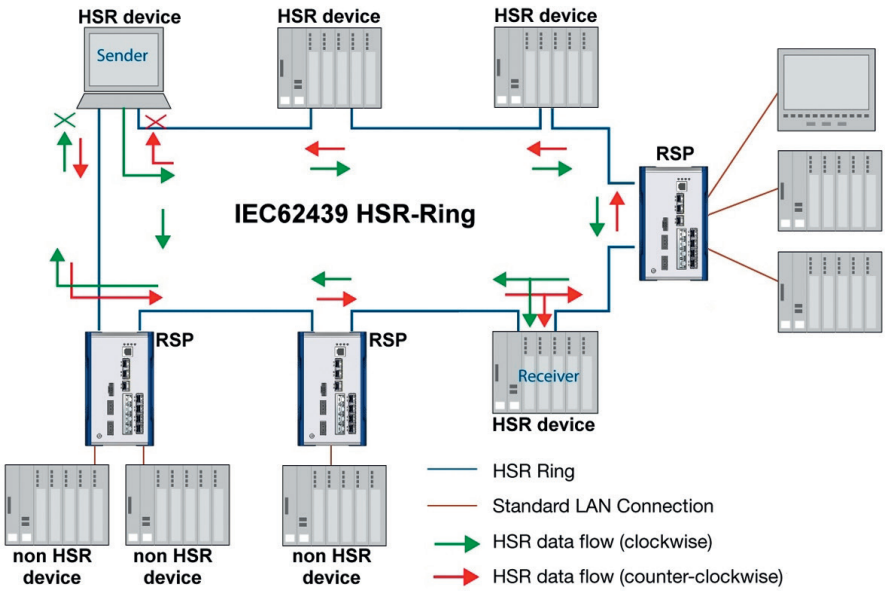


Figure 4 – Duplicate data packets are transmitted simultaneously in both directions

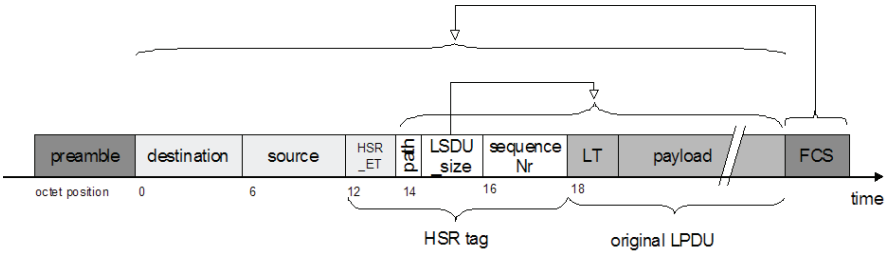


Figure 5 – HSR frame format with no VLAN tag (extract from IEC 62439 3)

In contrast to PRP, it is not possible to integrate SAN nodes directly into an HSR network without breaking the ring: a SAN lacks the second network interface necessary for a closed ring. This is one reason why SANs can be connected to HSR networks only via redundancy boxes. The second reason is the encapsulation of the network traffic on the ring effected by the HSR header. Unlike with PRP, this prevents ordinary network nodes from participating in the HSR traffic. While SAN nodes interpret the PRP RCTs as padding, this is not possible for the HSR tag: its position in the frame means that it is always interpreted as valid layer 2 frame information, and this prevents SAN nodes from correctly reading out the frame's user data.

Because some HSR devices may need to communicate with a management station or notebook for purposes of configuration and diagnostics, HSR connections will temporarily accept devices that transmit standard Ethernet frames, even on ringports. In this case the HSR connections communicate without HSR header encapsulation, although this traffic is not passed to the HSR network – it merely provides bidirectional communications between the configuring management station on an HSR port and the HSR device.

Normal HSR communications is not restarted until the ring has been closed. Couplings between two HSR rings are always implemented by means of two ring coupling elements, known as QuadBoxes. These facilitate a coupling between two HSR with no single point of failure (see Figure 6)

As regards switchover times, HSR behaves just like PRP: by sending duplicate frames from both the ports of an HSR connection, in the event of a failure one frame will still be transmitted via whichever network path is still intact.

This means that the redundancy again functions with zero switchover time and, unlike PRP, does not require two parallel networks.

An HSR network, however, always has the form of a ring, or a structure of coupled rings, which means that it is less flexible than PRP at the installation stage.

The duplicate transmission of frames in both directions also means that effectively only 50% of the network bandwidth is available for data traffic.

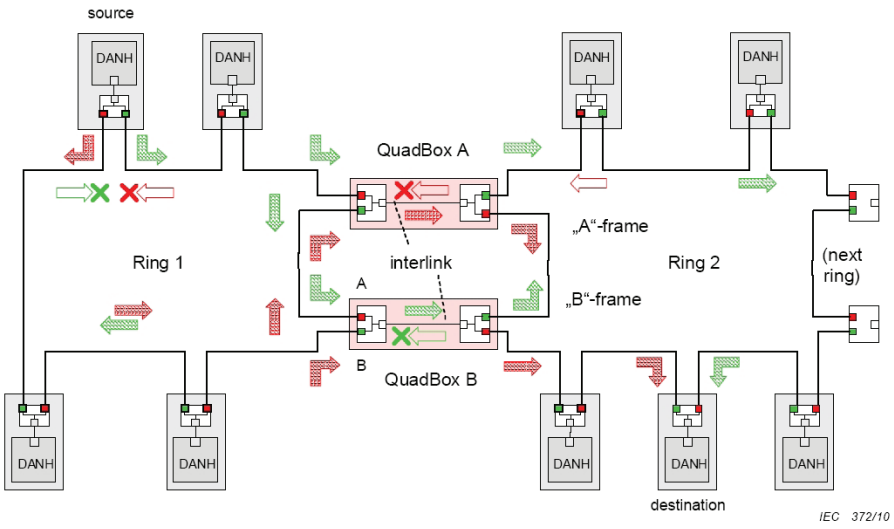


Figure 6 – Coupled HSR rings (extract from IEC 62439 3)

5. Summary

In practice, there is no perfect network topology nor perfect media redundancy protocol that precisely covers all applications and requirements.
The right choice of topology and protocol will always depend on additional factors, such as the physical installation requirements and/or the switchover times demanded by the application.
As an overview, the following table summarizes the protocols and principal parameters of the redundancy technologies covered in this white paper.

The current state of Ethernet technology is well able to fulfill the requirements of the most demanding applications. Provided the correct technology is identified during the planning phase of a communications network it is possible to minimize project risks at an early stage. This white paper constitutes an initial step towards identifying the most suitable technology. For additional discussions, service and advice please contact the consulting team at Belden's HiCom Center, who will be glad to assist you with advice and support in order to create a tailored solution for your individual needs [5].

Protocol	Topology	Max. devices	Worst-case reconfiguration time	Normal-case reconfiguration time
RSTP (IEEE 802.1D-2004)	Ring	40	Over 2s for loss of more than one BPDU	Depends on the RSTP implementation and the number of switches in the ring. Typically between 100ms and 200ms for 40 devices
RSTP (IEEE 802.1D-2004)	Any	Any	>2s for loss of more than one BPDU	Difficult to estimate, calls for detailed analysis of the individual network.
MRP (IEC 62439-2)	Ring	50	500ms, 200ms, 30ms, 10ms (depending on the supported parameter set)	Ca. 200ms, 60ms, 15ms, <10ms (depending on the supported parameter set)
PRP (IEC 62439-3)	Double, any	Any	0ms	0ms
HSR (IEC 62439-3)	(coupled) rings	512	0ms	0ms

6. References

[1] Hubert Kirrmann – Fault tolerant computing in industrial automation (http://lamspeople.epfl.ch/kirrmann/Pubs/FT_Tutorial_HK_050418.pdf)
[2] IEEE 802.1AX-2008 (http://standards.ieee.org/getieee802/download/802.1AX-2008.pdf)
[3] IEEE 802.1D-2004 (http://standards.ieee.org/getieee802/download/802.1D-2004.pdf)
[4] IEEE 802.1Q-2005/cor1-2008 (http://standards.ieee.org/getieee802/download/802.1Q-2005_Cor1-2008.pdf)
[5] Hirschmann Service und Support(http://www.beldensolutions.com/de/Service/index.phtml)

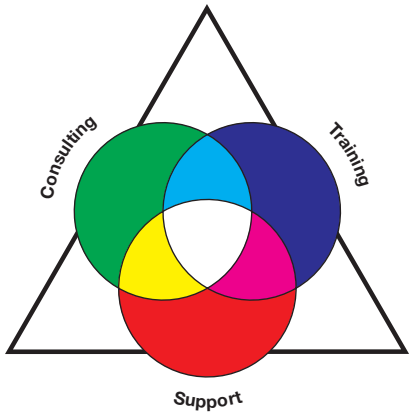
Annex:
Further Support

Technical Questions and Training Courses

In the event of technical queries, please contact your local Hirschmann™ distributor or Hirschmann™ office. You can find the addresses of our distributors on the Internet: www.hirschmann.com.

Our support line is also at your disposal:
Tel. +49 1805 14-1538
Fax +49 7127 14-1551

The current training courses to technology and products can be found under http://www.hicomcenter.com.



Belden Competence Center

In the long term, excellent products alone do not guarantee a successful customer relationship. Only comprehensive service makes a difference worldwide. In the current global competition scenario, the Belden Competence Center is ahead of its competitors on three counts with its complete range of innovative services:

- Consulting incorporates comprehensive technical advice, from system evaluation through network planning to project planing.
- Training offers you an introduction to the basics, product briefing and user training with certification.
- Support ranges from the first installation through the standby service to maintenance concepts.

With the Belden Competence Center, you have decided against making any compromises. Our client-customized package leaves you free to choose the service components you want to use. Internet: http://www.hicomcenter.com.

© 2014, Belden Inc.