

# **SERVICE PROVIDER**

# Achieving Superior Resiliency and Fast Convergence with Brocade Metro Ring Protocol (MRP)

Service providers are looking for new ways to produce incremental revenue by providing carrier-grade Ethernet services, business VPNs, and Internet connectivity. The Brocade MRP is specifically designed for metro rings as an alternative to Spanning Tree Protocol (STP), and it offers fault detection, isolation, and failover in just a few milliseconds.

**BROCADE** 

# **CONTENTS**

Contents	2
Introduction	
Benefits of MRP and VSRP	3
Ring Topologies	4
Basics of MRP Technology	5
Fault Detection and Failover	5
Support for Overlapping Rings with MRP-II	6
Detecting Failures in Overlapping Rings	8
Transitions When a Ring Is Healed	9
Efficient Ring Utilization	9
Integration with an MPLS Core	9
Scalability of a Ring	9
MRP Support in Brocade Products	10
Summary	10

INTRODUCTION

Service providers are looking for new ways to produce incremental revenue by providing carrier-grade Ethernet services, business VPNs, and Internet connectivity. The challenges in bringing these services to market are to minimize upfront investment and still maintain a high level of service in the infrastructure.

Using a Layer 2 architecture at the edge of a metro network provides a simple, low-cost solution. Building ring networks is a proven method to cost-effectively maximize the fiber layout so as to reach a large number of sites within a target geography, while still providing inherent redundancy. However, such ring topologies raise the question of the appropriate solution to use that meets scalability, availability, and efficient link utilization objectives. While traditional Spanning Tree Protocol (STP) provides a loop-free environment in arbitrary topologies, it can take up to 45 seconds to converge in the event of failure within a network. Rapid Spanning Tree Protocol (RSTP, IEEE 802.1w) provides improvements, lowering convergence time in the order of a second, but that is still not acceptable when offering carrier-grade Ethernet services.

The Layer 2 metro solution from Brocade includes two proven protocols designed to offer an alternative to Spanning Tree-based designs:

- Brocade® Metro Ring Protocol (MRP)¹ is specifically designed for metro rings as an alternative to Spanning Tree and offers very fast fault detection, isolation, and failover in just a few milliseconds.
   Brocade MRP provides protection similar to that promised by Resilient Packet Ring (RPR) but requires no special Ethernet PHYs. As a result, MRP can be used to meet the same goals at a dramatically different price point.
- Virtual Switch Redundancy Protocol (VSRP) is a "sister" technology to MRP, with the same goal—an
  alternative to STP with subsecond failover, yet specifically designed for mesh topologies. Combining
  these two protocols can provide an extremely robust network design.

This document explains the operation and benefits of the Brocade Metro Ring Protocol.

## **BENEFITS OF MRP AND VSRP**

A protocol called Ethernet Ring Protection (G.8032) was developed by the International Telecommunications Union (ITU) to address similar goals. Compared to G.8032, the Brocade MRP and VSRP protocols offer the following benefits:

- G.8032 does not support overlapping rings, that is, rings that have shared links. In large ring networks,
  it is optimal to build rings with shared links. In contrast, Brocade MRP-II supports overlapping rings.
- G.8032 works specifically in ring environments and does not work in mesh designs. Typical networks
  have a mix of mesh and ring-based topologies. In contrast, Brocade VSRP supports mesh topologies.
  Additionally, VSRP and MRP have been integrated to provide efficient failover.
- G.8032 requires specialized hardware to process protocol messages, particularly when operating at
  very high speeds. In contrast, the MRP protocol can be implemented with hardware-based protocol
  forwarding on any class of Ethernet devices, thereby capitalizing on the economies of scale offered by
  mass-market Ethernet.

<sup>&</sup>lt;sup>1</sup> This document uses the abbreviation MRP to refer to the Brocade Metro Ring Protocol. The reference to MRP must not be confused with the term "Multiple Registration Protocol," which is specified in IEEE 802.1ak. The two protocols are totally independent of one another and have no commonality whatsoever.

Both MRP and VSRP have been proven in very large Layer 2 network deployments for several years with a variety of Brocade products.

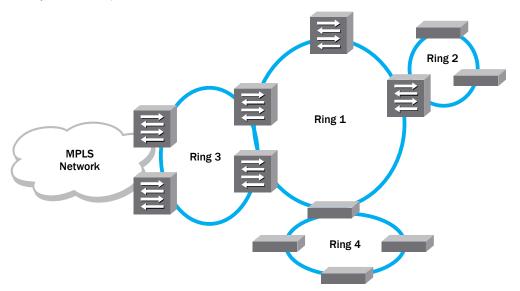


Figure 1. A typical metro network with multiple rings

Figure 1 shows a traditional metro ring topology, which may utilize existing SONET infrastructure, Gigabit Ethernet (GbE), and 10 GbE in overlapping or non-overlapping rings.

## **RING TOPOLOGIES**

There are three basic ring topologies, as shown in Figure 2, which cover all metro ring designs:

- Single ring. A ring with no direct connection to other rings
- Non-overlapping rings. Rings that connect to a single switch but do not share any links among them
- Overlapping (or shared) rings. Rings that overlap with shared links, that is, links that impact the fate of multiple rings

Brocade MRP is flexible enough to handle all three of these ring topologies.

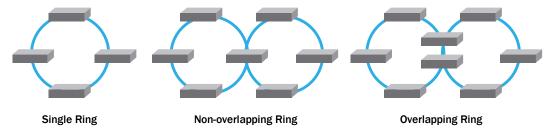


Figure 2. Types of rings

#### **BASICS OF MRP TECHNOLOGY**

The primary goal of MRP is to provide a loop-free topology. In order to accomplish this, one switch in each ring, along with its ring interfaces, is configured as the "Master" of the ring. MRP creates a loop-free ring topology by declaring one of the Master's ring ports to be in the "Blocking" state. When a port is in the Blocking state, it receives only MRP protocol packets and discards any other data packets. In the event of link or node failure in the ring, MRP changes the Master's Blocking port to the "Forwarding" state, thereby allowing traffic to avoid the failed link.

Once configured for MRP, the Master switch monitors the health of the ring by sending a specific protocol packet called the Ring Hello Packet (RHP) on its forwarding interface at a specific interval. This interval is called the "hello time," and it is configurable in 100-ms increments. The interface of a node on which RHP packets are forwarded is called the primary interface, and the interface on which RHP packets are received is called the secondary interface. The non-Master switches in the ring forward the received RHP packets in hardware to its primary interface. If the ring is healthy, the Master receives the RHP on its secondary interface. Only the secondary interface of the Master node is in the Blocking state in a healthy ring.

#### **FAULT DETECTION AND FAILOVER**

When a fault such as link failure occurs in the ring, the node adjacent to the failed link detects and sends a special type of RHP packet called an "Alarm RHP" to report a fault in a network. On receiving the Alarm RHP packet, the MRP Master transitions its secondary interface from Blocking to Forwarding, thereby immediately healing the ring. Subsequently, it sends three Topology Change Notification (TCN) messages to ring members, alerting the nodes to flush their MAC address tables. The sequence of operations for a link failure is depicted in Figure 3.

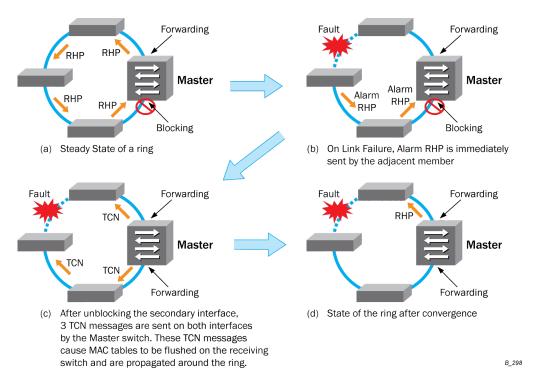


Figure 3. Sequence of operations in MRP during link failure

MRP is also resilient to detect node failures such as a failed node—or even a rare event such as a data plane failure—on one of its member nodes. In such a scenario, the RHP packets sent by the Master node are no longer received on the Master's secondary interface. If the Master does not receive the RHP packet on its secondary interface, it waits for a certain period of time, called the "dead timer." Upon expiry of the dead timer, the secondary interface on the Master switch transitions from the "Blocked" to the "Preforwarding" state. In the Preforwarding state, the blocked interface listens only for RHP packets. The Preforwarding delay is configurable as a multiple of the hello time and must be at least twice the hello time. The Preforwarding state is necessary to prevent loops and ring oscillations during failure/recovery.

On completion of the Preforwarding delay time, the secondary interface on the Master switch transitions from a Preforwarding to a Forwarding state. As in the case of link failures, the Master Switch subsequently sends three TCN messages on all the ring interfaces, triggering a flush of the forwarding tables on MRP member nodes. All MRP nodes in the ring now learn the new topology.

# SUPPORT FOR OVERLAPPING RINGS WITH MRP-II

A common practice in building out large metro networks is to utilize the fiber plant efficiently by allowing overlapping links between different rings. The Brocade MRP-II implementation allows rings to overlap and use a shared interface among those rings. In such a scenario, every switch maintains state for each of the rings it is a part of, in order to keep track of the appropriate interface that needs to forward the RHPs.

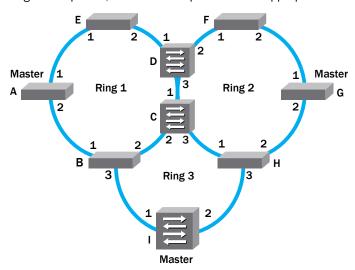


Figure 4. An example of a network with overlapping rings

In order to support overlapping rings, MRP-II uses the concepts of ring priority and shared interfaces. *Ring priority is based on the ID of the ring itself*, with higher numerical values denoting a lower priority. For example, in Figure 4, Ring 1 has the highest priority and Ring 3 has the lowest priority. A shared interface is an interface that is connected to multiple rings. For example, Interface 2 on Switch B is a shared interface, since it is part of both Ring 1 and Ring 3. Similarly, Interface 3 on Switch C is a shared interface, because it is part of both Ring 2 and Ring 3.

The key characteristic of MRP-II is that a node sends RHP packets *only* to all interfaces on higher priority rings in the same VLAN. This behavior ensures that RHP packets from higher priority rings never traverse a lower priority ring. Figure 5 demonstrates the flow of RHP packets for the network shown in Figure 4.

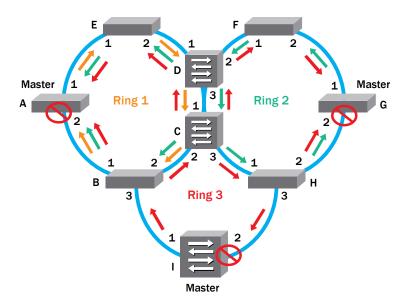


Figure 5. Forwarding of RHP packets in the three overlapping rings

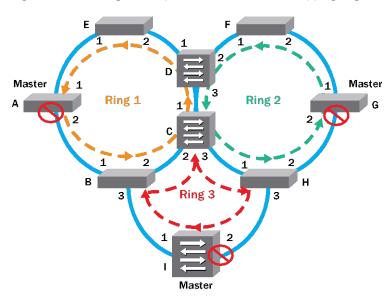


Figure 6. Flow of RHP packets to complete the ring.

As shown in Figure 5:

- RHP packets from Ring 1 do not flow on Ring 2.
- RHP packets from Ring 2 flow on Ring 1 but do not flow on Ring 3.
- RHP packets from Ring 2 flow on Ring 1 because Ring 1 is a higher priority ring.
- On the Master nodes, RHP packets are not forwarded out a blocked interface. Further, RHP packets received on a blocked interface for which that node is not the Master are also ignored. Thus, RHP packets for Ring 2 (green) and Ring 3 (red) received by Node A on Interface 1 are not forwarded to Interface 2. Similarly, RHP packets for Ring 2 (green) and Ring 3 (red) received by Node A on Interface 2 are ignored.

The actual flow of RHP packets to complete each of the rings is shown in Figure 6.

## **DETECTING FAILURES IN OVERLAPPING RINGS**

When a failure occurs on a link connected to a regular (non-shared) interface, the same procedure is used as in the case of MRP. On the other hand, if the failure happens on a shared interface, then the ordered, cascaded RHP flow already taking place, from lower priority rings over higher priority rings, will ensure a loop free topology. This is demonstrated in Figure 7 and Figure 8. Switch A transitions Interface 2 into the Preforwarding state when it gets an Alarm RHP for Ring 1. Alternatively, if Switch A stops receiving RHP packets for Ring 1, Interface 2 on Switch A transitions from the Blocking -> Preforwarding state. In the Preforwarding state, there is no data forwarding on Interface 2 of Switch A. However, RHP packets from Ring 2 will traverse Ring 1 completely then arrive at Switch G (Master for Ring 2). Thus, Interface 2 on Switch G continues to be blocked. After the Preforwarding interval, Switch A transitions Interface 2 into the forwarding state for actual data forwarding. Hence, throughout the reconvergence process, transient forwarding loops are avoided.

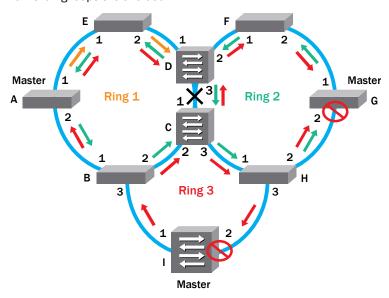


Figure 7. Forwarding of RHP packets when the shared interface between Switches C and D fails

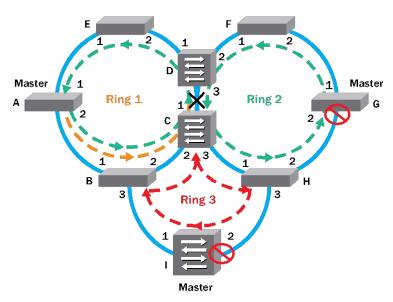


Figure 8. Flow of RHP packets to complete the rings when a shared interface fails

TRANSITIONS WHEN A RING IS HEALED

When a failed link recovers, the nodes adjacent to the recovered link place the interface in the Preforwarding state. In this state, only RHP packets are forwarded. The periodic RHP packets sent by the Master are thus received back by the Master switch on its secondary interface. This notifies the Master switch that the ring has potentially healed. At this stage, the secondary interface on the Master switch transitions to the Blocking state. Subsequently, three TCN messages are sent by the MRP Master on its forwarding interface of that ring to notify ring members of a topology change. The TCN message is used as a signal to flush the MAC tables on the member nodes for the affected VLANs. The Master node also sends a special RHP packet notifying the member nodes to transition to the Forwarding state. Upon receipt of this RHP packet (or upon expiry of the Preforwarding time), the nodes adjacent to the recovered link transition that interface from the Preforwarding to the Forwarding state.

#### **EFFICIENT RING UTILIZATION**

Brocade routers and switches offer two mechanisms to further improve the efficiency of the ring. When the ring's Master node blocks a port, that segment of the ring is not utilized (except for control traffic), wasting valuable resources. MRP solves this problem by allowing multiple logical rings over the same physical links.

The second optimization is the use of topology groups. In networks with a large number of VLANs, the overhead to run a separate instance of a control protocol could increase load on the CPU. To prevent high CPU load, Brocade switches and routers use topology groups that share bridge topologies among different VLANs. Each topology group can have independent interface states, so one topology group can be Blocking on an interface while another is Forwarding. A topology group controls the state of specific switch ports in use by VLANs or group of VLANs.

Using topology groups is an efficient method of decreasing CPU load on switches by applying topology information learned on a "control" VLAN to other members that are in the same topology group as the control VLAN.

The above two optimizations may also be combined to get even further benefits. Each topology group can have a unique MRP Master switch. When the Master blocks a port, it does so for only the specified control VLAN (and all VLANs controlled by that topology group). Figure 9 shows links Blocked for certain VLANs while Forwarding for others. Combining these two thus provides efficient link utilization.

#### INTEGRATION WITH AN MPLS CORE

One of the most proven and resilient methods of deploying carrier-grade Ethernet service is over an MPLS core. Customer VLANs in an access ring may thus be carried over an MPLS core by mapping them to a VPLS instance. By using MRP in access rings that terminate on the Provider Edge router, a VPLS service can be offered with end-to-end resiliency that seamlessly integrates with the resiliency offered by an MPLS network.

# **SCALABILITY OF A RING**

There is theoretically no upper limit to the number of nodes that may be present in a ring that is running MRP. In practice, the size of a ring is limited by the *ring latency*. Ring latency is the time taken by a RHP packet to travel around the ring and come back to the MRP Master node. The ring latency comprises of the per-node forwarding latency and the sum of the latencies on the physical link between each member node in a ring. For proper MRP operation, it is recommended that ring latency be less than the hello interval. In no case must the ring latency be greater than the hello interval + dead timer interval. Rings containing 16 nodes or greater can be achieved and have been successfully implemented in production networks. MRP is also flexible to allow ring nodes to be interconnected by a Link Aggregation Group (LAG), should you desire additional capacity in the ring.

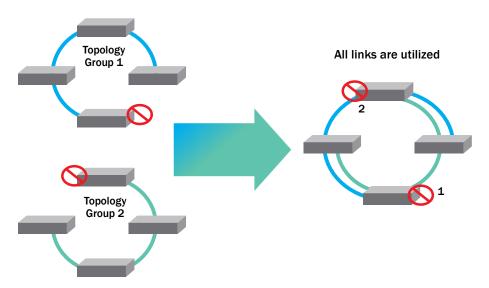


Figure 9. Increasing ring utilization using topology groups

# MRP SUPPORT IN BROCADE PRODUCTS

MRP is currently supported on a broad range of Brocade switches and routers, in both fixed form factor and modular chassis products. For details on MRP support in Brocade service provider products, visit: http://www.brocade.com/dotcom/products-solutions/products/ethernet-switches-routers/service-provider/index.page?

#### **SUMMARY**

The Brocade MRP offers an alternative to Spanning Tree and provides super resiliency in ring networks. Its support for different types of ring topologies has made the Brocade MRP a popular choice for providers who want to balance advanced resiliency requirements with a cost-effective loop-prevention mechanism in Layer 2 networks. Further, the Brocade MRP integration with VSRP, VPLS, topology groups, and VLAN groups maximizes utilization of the ring and allows end-to-end resiliency to be achieved when delivering a carrier-grade Ethernet service.

© 2012 Brocade Communications Systems, Inc. All Rights Reserved. 09/12 GA-TB-238-01

ADX, Brocade, Brocade Assurance, Brocade One, the B-wing symbol, DCX, Fabric OS, ICX, MLX, SAN Health, VCS, and VDX are registered trademarks, and AnylO, HyperEdge, MyBrocade, NET Health, OpenScript, and The Effortless Network are trademarks of Brocade Communications Systems, Inc., in the United States and/or in other countries. Other brands, products, or service names mentioned may be trademarks of their respective owners.

Notice: This document is for informational purposes only and does not set forth any warranty, expressed or implied, concerning any equipment, equipment feature, or service offered or to be offered by Brocade. Brocade reserves the right to make changes to this document at any time, without notice, and assumes no responsibility for its use. This informational document describes features that may not be currently available. Contact a Brocade sales office for information on feature and product availability. Export of technical data contained in this document may require an export license from the United States government.