Cisco Resilient Ethernet Protocol

Introduction

The adoption of Carrier Ethernet is causing the creation of large Layer 2 domains that require fast and predictable convergence. In particular, Service providers require fast convergence to support video and voice services for their triple-play deployments. Moreover, this mechanism for resilient failovers must be flexible and it must support complex ring topologies. It also should support fast convergence when scaling the number of VLANs and MAC addresses. Consequently, Cisco® has designed a new Resilient Ethernet Protocol (REP) that meets these requirements for fast and predictable convergence in Layer 2 ring topologies.

Technology Overview

Cisco Resilient Ethernet Protocol is a new technology implemented on Cisco Carrier Ethernet switches and intelligent service edge routers. This software enhancement for Cisco Carrier Ethernet platforms extends network resiliency across Cisco IP Next-Generation Network (NGN) Carrier Ethernet Design. Requiring no hardware upgrades, REP is designed to provide network and application convergence within 50 ms. In some scenarios, the network convergence times may increase to within 250 ms, but a 250-ms convergence time is still expected to have limited or no discernable effect on most network applications. REP is a segment protocol that integrates easily into existing Carrier Ethernet networks. It does not intend to replace the Spanning Tree Protocol, but allows network architects to limit the scope of Spanning Tree Protocol domains. Since Cisco REP can also notify the Spanning Tree Protocol about potential topology changes, it allows for interoperability with Spanning Tree. Ideally, REP can be positioned as a migration strategy from legacy spanning tree domains.

Cisco REP is easy to configure and manage, using tools such as topology archiving to simplify network management. Its preemption mechanism also makes the network more predictable. Because REP is a distributed and secure protocol, it does not rely on a master node controlling the status of the ring. Hence failures can be detected locally either through loss of signal (LOS) or loss of neighbor adjacency. Any REP port can initiate a switchover as long as it has acquired the secure key to unblock the alternate port. By default, REP elects an alternate port unless the administrator defines a preferred port. For optimal bandwidth usage and for traffic engineering, REP supports load balancing per group of VLANs.

Definition of REP Terms

Cisco REP is a segment protocol; a REP segment is a chain of ports connected to each other and configured with the same segment ID. Each end of a segment terminates on an edge switch. The port where the segment terminates is called the edge port. Figure 1 illustrates a REP segment. This basic element makes REP extremely flexible in the way you can plug this topology entity into existing topologies, which can include ring, dual home, and hub & spoke designs to name a few.

Figure 1. A REP Segment

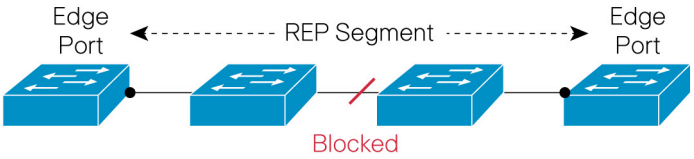


Figure 2 shows how REP wraps into a ring topology. Note that each node in the segment has exactly 2 REP-enabled ports.

Figure 2. REP Wrapped into a Ring Topology

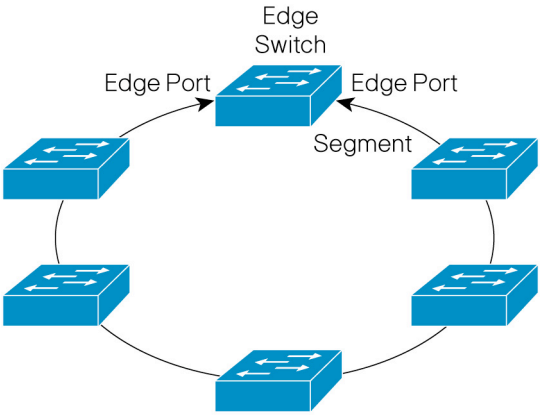
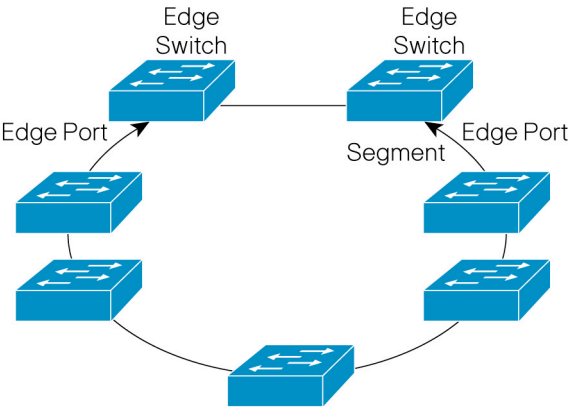


Figure 3 illustrates how a REP segment can terminate on different edge switches. Supporting this topology is simplified by the segment element; typical ring resilient protocols require redundant master node functions to accomplish the same topology. Also note that the link between the two edge switches could be running Spanning Tree.

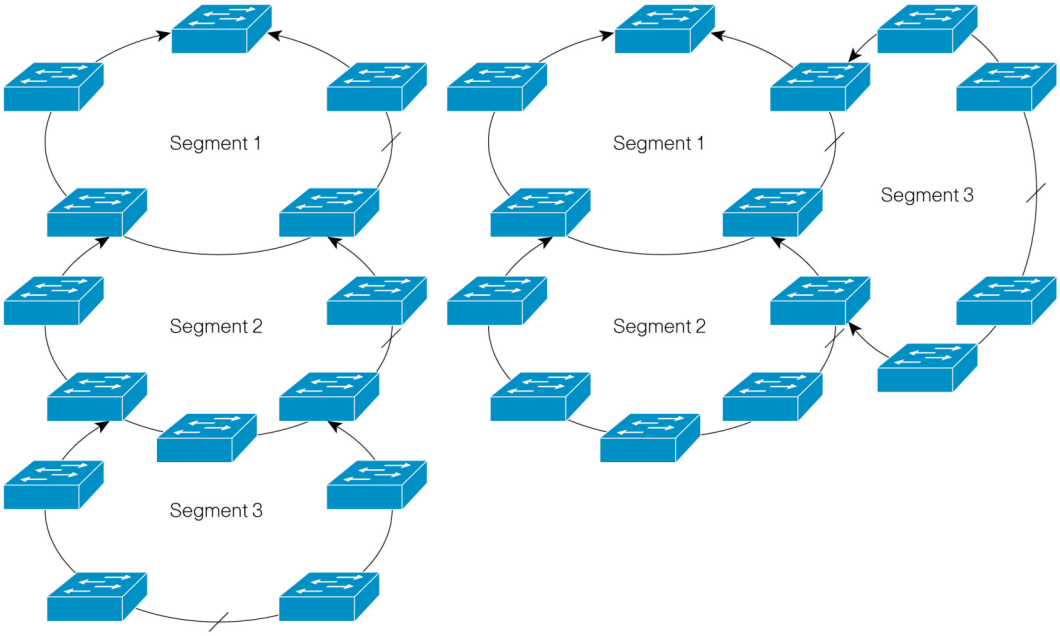
Figure 3. A REP Segment Terminating on Two Edge Switches



Cisco REP: Support for More Complex Topologies

Cisco REP provides increased flexibility by supporting a variety of topologies. Figure 4 illustrates how REP supports complex ring topologies.

Figure 4. REP for More Complex Ring Topologies

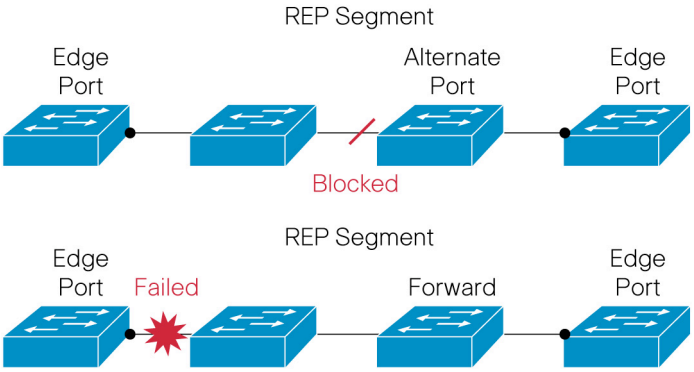


Typical ring resiliency protocols require rings to be closed, whereas Cisco REP supports rings that are open or closed. This new protocol also supports topologies with redundant aggregation devices. As shown in Figure 4, each segment is identified with a unique segment ID. Note that each REP port can support only one segment ID, because each switch is configured with a maximum of 2 REP ports with the same segment ID.

Cisco REP Operation

With REP at least one port is always blocked in any given segment, that is, the alternate port. The blocked port helps ensure that the traffic within the segment is loop-free by requiring traffic flow to exit only one of the edge ports, and not both. So when a failure occurs in the segment, REP opens the alternate port so traffic can reach the edge of the segment (Figure 5).

Figure 5. Cisco REP Basic Operation



Cisco REP Fault Detection

Cisco REP relies primarily on LOS to detect a link failure, and can always learn the location of the failure within the ring. When a failure occurs, the failed ports immediately send link failure notifications to all REP peers. The failure notification has two purposes:

- Instruct the alternate port to immediately unblock because the segment is broken.
- Flush MAC entries on all REP ports within the segment.

A REP node maintains neighbor adjacencies and continuously exchanges hello packets with its neighbors. In scenarios where LOS is not detected, the loss of a REP adjacency also triggers a switchover. Neighbor adjacency awareness is unique to REP and has advantages over alternate polling mechanisms that require centralized management from a master node. Note that the Unidirectional Link Detection Protocol (UDLD) can be enabled on REP interfaces to detect unidirectional failures.

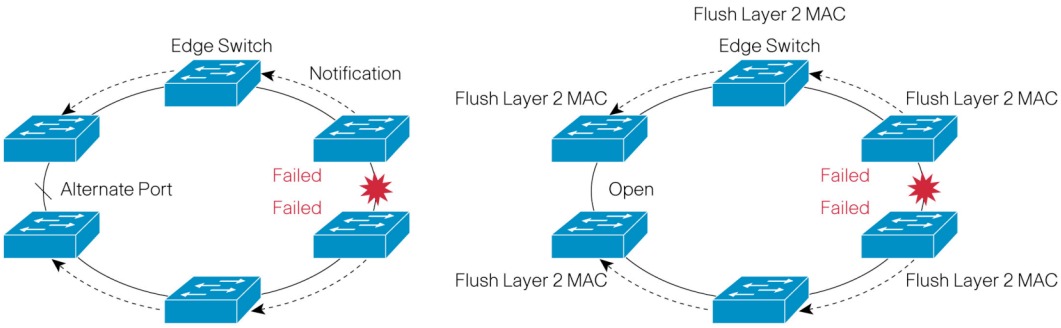
Cisco REP Failure Notification

Fast failure notification is a critical requirement to accomplish fast convergence. To ensure reliable and fast notification, Cisco REP propagates the notifications using the following two methods:

- Fast notification: Using a Cisco Multicast address, the notification is forwarded in hardware so that each node in the segment is notified immediately without software involvement from any node.
- Reliable notification: Distributed through the REP Adjacency Protocol, the notification is retransmitted if lost. The protocol uses sequence numbering and relies on packet acknowledgment.

Upon receiving the notification, each REP node flushes MAC address entries learned on these REP ports, and the alternate port then begins forwarding traffic. Because the notification is sent through a Cisco reserved Multicast address, the MAC addresses flushing can proceed in parallel on each REP node (Figure 6).

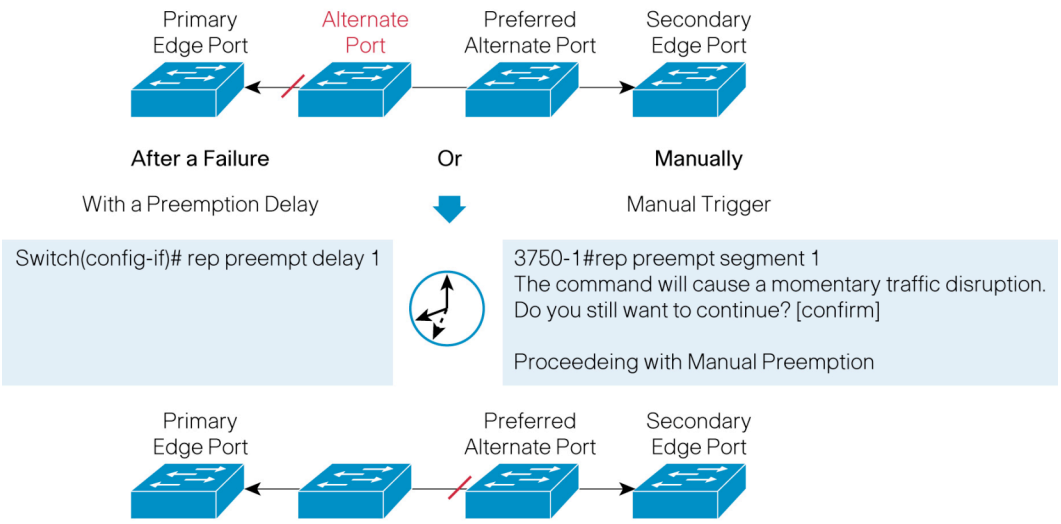
Figure 6. Cisco REP Link Fault Notification



Cisco REP Fault Restoration and Preemption

Preemption is the mechanism that allows a REP segment to return to a well-known state after a failure. By default, the REP preemption mechanism is not active, meaning it avoids disrupting the traffic after fault restoration. It avoids traffic disruption by blocking one of the two restored ports so that bidirectional traffic is not affected and backhauling of unidirectional traffic is avoided. However, with preemption enabled, REP transitions to its well-known state after the preemption delay expires. Here the administrator should define the preferred alternate port for that well-known state. In summary, preemption can be triggered either manually or by using a preemption delay timer (Figure 7).

Figure 7. Cisco REP Preemption

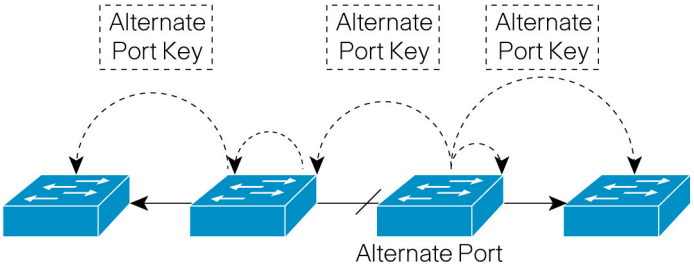


Cisco REP: Distributed and Secure

Cisco REP is a distributed and secure protocol that does not rely on a master node monitoring the health of the ring. The primary edge port is responsible only for initiating topology collection and the preemption process. Failure can be detected locally either through LOS or loss of neighbor adjacency. Any REP port can initiate a switchover as long as it has acquired a secure key to unblock the alternate port. The secure key consists of a 9-byte length word that identifies each port. It is a combination of the port ID and a random number generated when the port activates. The alternate port key is secure because it is distributed only within the segment.

The REP alternate port generates and distributes its key to all other ports within the segment (Figure 8). Each port on the segment can use that key to unblock the alternate port. With this mechanism, users cannot unblock the *alternate* port unless they learn the key. This mechanism protects against potential security attacks; it also avoids problems with overlapping segment IDs. (Note that with 1024 segment IDs available, overlapping most likely will not occur, but misconfiguration could lead to such a scenario.)

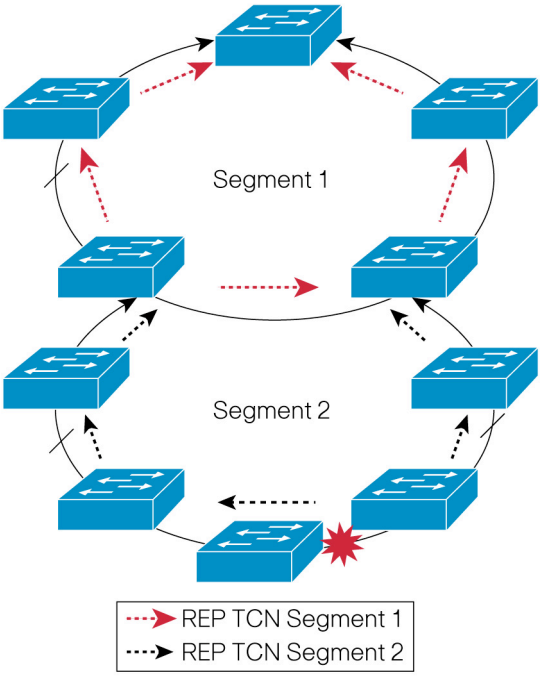
Figure 8. Alternate Port Key Distribution



Topology Change Notification

Topology change notification (TCN) is used within the segment to notify REP neighbors of topology changes. At the edge of the segment, REP can propagate the notification to Spanning Tree Protocol or other REP segments. Figure 9 illustrates a scenario in which REP segments are configured on multiple rings. Segment 2 (S2) is configured to send TCN notification to segment 1 (S1), so that S1 flushes MAC entries of the host devices, thus avoiding black holing of unidirectional traffic.

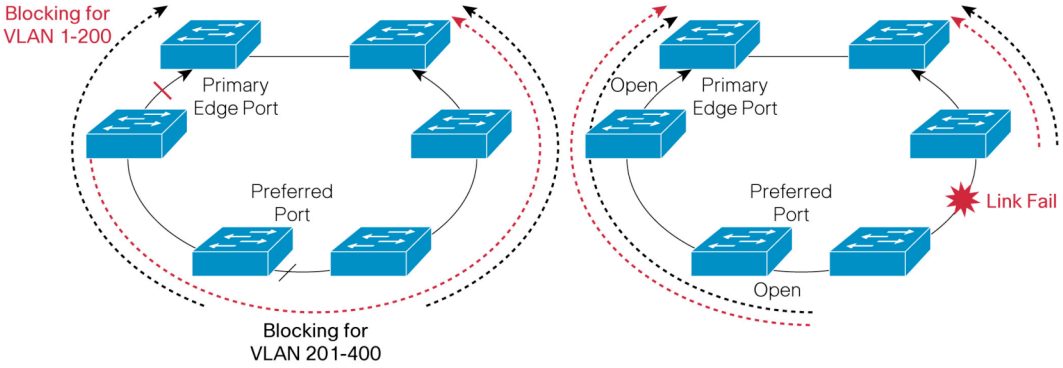
Figure 9. Topology Change Notification from One REP Segment to Another



VLAN Load Balancing for Optimal Bandwidth Usage in Rings

For optimal bandwidth usage in the ring, Cisco REP can load balance traffic across the ring. VLANs can be grouped into two instances. The first set of VLANs can be blocked at the primary edge port, and the second set of VLANs can be blocked at an alternate, predefined location. VLAN load balancing is centrally configurable on the primary edge switch, making VLAN traffic management more scalable. Figure 10 illustrates the load-balancing capability of REP.

Figure 10. VLAN Load Balancing



When a link failure occurs in the segment, both blocking ports become *open*, as shown in Figure 10. When the broken link is restored and the preemption delay has expired, REP preempts back to its load-balancing scheme.

Ease of Configuration and Management

Cisco REP configuration requires very few steps. With the preferred alternate port and preemption mechanism, the topology is well-known and the management simplified. The toolset includes a topology reporting tool that reports current and archived topology (Figure 11).

Figure 11. Output of show rep topology Command

```
3750-ME# show rep topology
REP Segment 1
BridgeName PortName Edge Role
-------------------------
3750-E Gi1/1/1 Pri      Open
3400-3 Gi0/2            Open
3400-3 Gi0/11           Open
3400-2 Gi0/2            Open
3400-2 Gi0/1            Open
3400-1 Gi0/2            Open
3400-1 Gi0/1            Alt
3750-E Gi1/1/2 Sec      Open
```

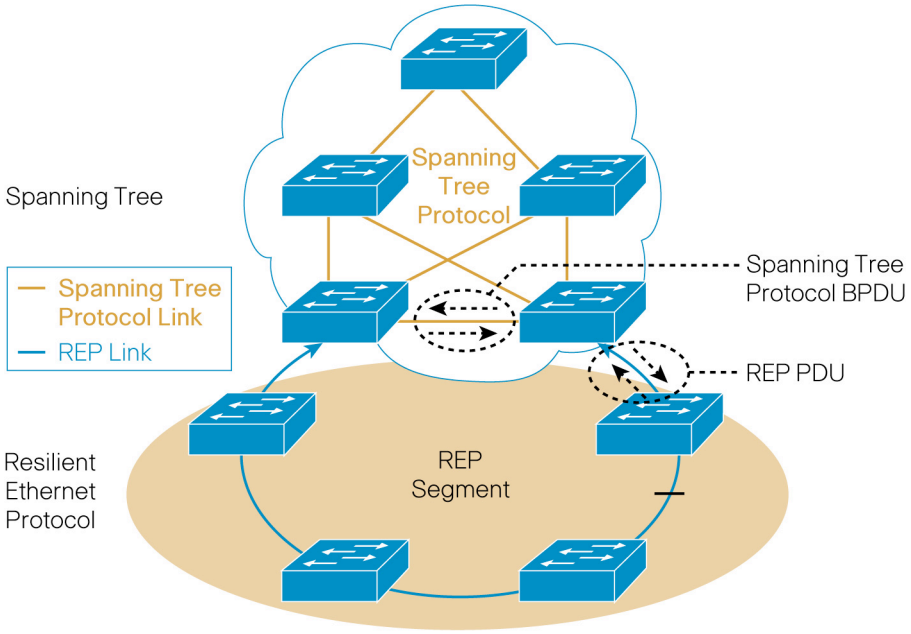
A Resilient Ethernet Protocol MIB is also available for SNMP management purposes.

Cisco REP and Spanning Tree Protocol

Cisco REP and Spanning Tree Protocol can coexist on the same switch, but not on the same interface; on the same interface, REP and Spanning Tree Protocol are mutually exclusive. If an interface is configured as a REP port, then Spanning Tree Protocol is disabled, and conversely. Although REP ports do not forward Spanning Tree Protocol bridge protocol data units (BPDUs),

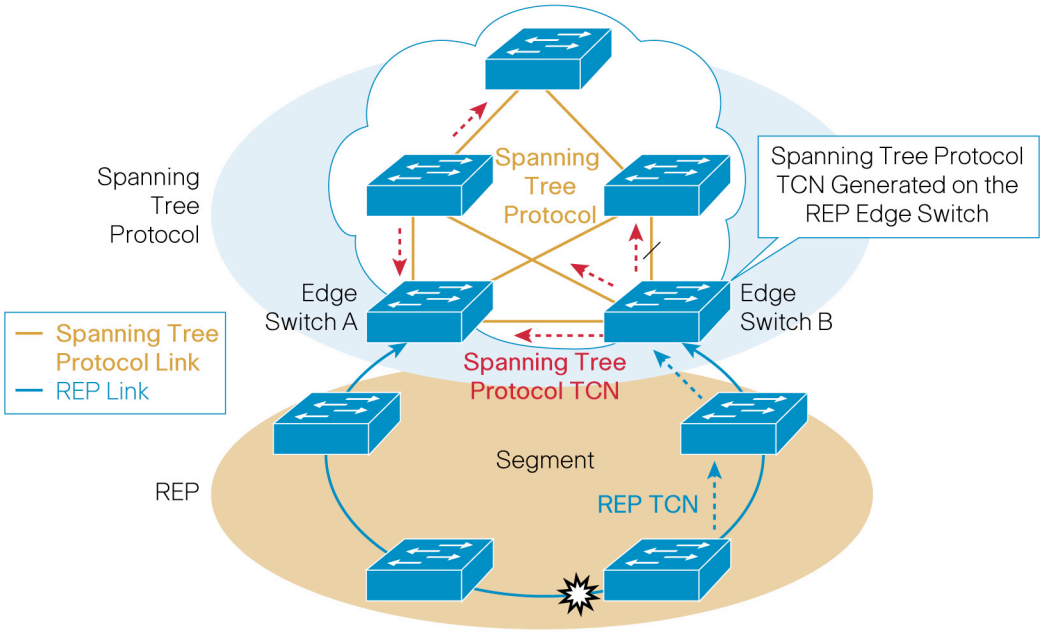
REP allows adjacent REP and Spanning Tree Protocol rings or domains to share a common link. This common link can be used for passing REP and STP date plane traffic, or for the Spanning Tree Protocol control plane traffic.

Figure 12. Enabling Spanning Tree Protocol and REP on Adjacent Domains



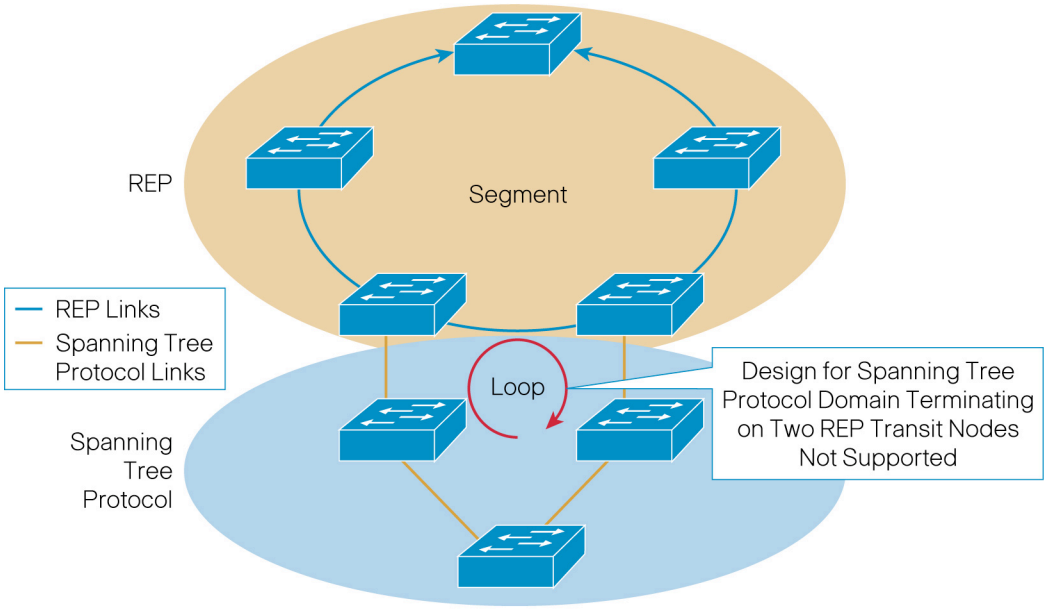
Since REP edge switches can be configured to propagate REP topology changes into the Spanning Tree domain, REP can notify the Spanning Tree Protocol about potential topology changes. Note that REP is able to generate Spanning Tree Topology Changes Notification (STCN). Figure 13 illustrates the REP topology changes notification being propagated in the Spanning Tree domain, which allows for interoperability between the two domains.

Figure 13. Topology Changes Notification propagated into Spanning Tree



Note that a design in which the Spanning Tree Protocol domain has dual connectivity to two REP transit nodes is not supported (Figure 14).

Figure 14. Cisco REP and Spanning Tree Protocol: A design which is not recommended



Conclusion

The Cisco Resilient Ethernet Protocol (REP) is designed to meet fast convergence requirements for Layer 2 domains. Ideally suited for ring configurations, it uses a native Ethernet element (the segment) that allows support for other topologies as well. REP is compatible and complementary with standard IEEE 802.1 spanning-tree protocols. In particular, TCNs outside the segments allow REP and Spanning Tree Protocol to operate in adjacent segments. Furthermore, REP is very simple to implement, and with features such as preemption and VLAN load balancing, it can be appropriate for service provider applications. Finally, REP will be available on a wide range of Cisco Carrier Ethernet switching and edge router platforms, extending consistent network recovery capabilities across the Cisco IP NGN Carrier Ethernet Design.

Americas Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
www.cisco.com
Tel: 408 526-4000
800 553-NETS (6387)
Fax: 408 527-0883

Asia Pacific Headquarters
Cisco Systems, Inc.
168 Robinson Road
#28-01 Capital Tower
Singapore 068912
www.cisco.com
Tel: +65 6317 7777
Fax: +65 6317 7799

Europe Headquarters
Cisco Systems International BV
Haarlerbergpark
Haarlerbergweg 13-19
1101 CH Amsterdam
The Netherlands
www-europe.cisco.com
Tel: +31 0 800 020 0791
Fax: +31 0 20 357 1100

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at **www.cisco.com/go/offices.**

Printed in USA

C11-427224-00 09/07