Extreme Standby Router Protocol and Virtual Routing Redundancy Protocol

Abstract: This paper will discuss two different protocols that increase network resiliency. Extreme Standby Router Protocol™ (ESRP) is a protocol created by Extreme Networks® to provide redundancy at both Layer 2 and Layer 3. Virtual Router Redundancy Protocol (VRRP) is a similar protocol, put forward by the IETF, that is designed specifically for Layer 3 redundancy in IP networks. VRRP used in conjunction with a Layer 2 loop prevention protocol such as Spanning Tree provides the functionality of ESRP.

Introduction

Higher standards of fault tolerance and resiliency are increasingly important when designing and implementing networks. How this is achieved will vary depending if the network is operating at Layer 2 or Layer 3.

This paper will discuss two different protocols that increase network resiliency. Extreme Standby Router Protocol (ESRP) is a protocol created by Extreme Networks to provide redundancy at both Layer 2 and Layer 3. Virtual Router Redundancy Protocol (VRRP) is a similar protocol, put forward by the IETF, that is designed specifically for Layer 3 redundancy in IP networks. VRRP used in conjunction with a Layer 2 loop prevention protocol such as Spanning Tree provides the functionality of ESRP.

ESRP and VRRP employ very similar concepts. Both protocols employ "Master-Slave" functionality, and provide redundant routing services to users. Only one Master device is active at any time, and a Slave device will take over the MAC address and IP address of the Master in the event of a failure. Both ESRP and VRRP provide redundancy for a "default gateway" service, while ESRP provides the additional capability of redundant Layer 2 switching.

The remainder of this whitepaper will delve into the details of where and why protocols such as ESRP and VRRP are needed, and how they increase network simplicity and resiliency.

Resiliency at Layer 2

Resiliency at Layer 2 is comparatively simple to achieve. All you have to do is ensure that there are multiple paths between different destinations. Of course, having multiple paths between network elements means that there will be both logical and physical loops in the topology, and Layer 2 loop prevention techniques must be employed to prevent traffic storms. Spanning Tree (802.1D), Rapid Spanning Tree (802.1w), or Extreme Automatic Protection Switching (EAPS) which can only be used in ring topologies, are all options. Time required for switchover is key—Spanning Tree takes up to 30 seconds to converge, and a 30 second outage is unacceptable. By contrast, EAPS offers convergence in sub-50 ms in most deployments.

In simple topologies where switches are connecting to multiple upstream devices—in case one of them fails—the aforementioned loop prevention protocols are overkill. A simplified approach can improve failover times. For this, Extreme Networks uses ESRP; the details of ESRP will be discussed later in this whitepaper.

Resiliency at Layer 3

Providing resiliency in Layer 3 topologies is somewhat more complicated than Layer 2. One method is to run a routing protocol such as RIP or Open Shortest Path First (OSPF) to the edge of the network. When an upstream router or path fails, the other routers will dynamically learn alternative routes. Unfortunately, there are a number of downsides to this:

- Inexpensive Layer 2 devices cannot be deployed at the edge
- Routing protocol convergence time increases with the number of routers
- Scalability is reduced
- Troubleshooting becomes more complex

As a result, many networks have been designed with Layer 2 devices to aggregate endstations, which utilize a statically configured gateway. These endstations are considered "dumb" because they can't dynamically change the gateway address if the upstream router becomes unavailable. Therefore, these endstations lose connectivity.

Both ESRP and VRRP are protocols where multiple routers "back each other up" in the event that the primary router fails. The routers communicate with each other through either the VRRP or ESRP protocol. One router is elected Master; if it fails, the backup router will assume the Master's IP and MAC addresses. This allows the failure to be transparent to the endstations.

Introduction to VRRP

VRRP works well in a network scenario where a single subnet feeds through a simple Layer 2 device and terminates on a router port—the traditional network topology prior to the popularization of Layer 2/Layer 3 switches.

When the Master router becomes unavailable, failover takes 3 seconds. The default advertisement interval is 1 second, and Master Down is declared after 3 advertisement intervals are missed. (Note that in a normal state, only the Master responds to ARPs and advertisements.) When a new router takes over as Master, an ARP request is sent out to each IP address on the VLAN—so that it can quickly learn all MAC addresses on the VLAN.

VRRP is strictly a Layer 3 protocol. In other words, some loop prevention protocol such as EAPS or 802.1D Spanning Tree or 802.1w Rapid Spanning Tree must be run in conjunction with VRRP to prevent loops at Layer 2. Running VRRP in conjunction with a Layer 2 loop prevention protocol potentially leads to a number of unexpected traffic patterns. As an example, assume the following topology:

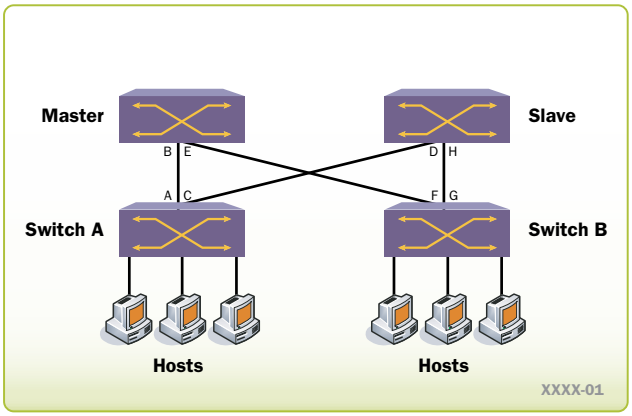

Figure 1

- The Layer 2 loop prevention protocol may put a link to the Master in blocking state. In the configuration above, there is a loop (ABEFGHDC). Suppose the Spanning Tree protocol blocks link AB. Hosts connected to Switch A must follow the path from Switch A to the Slave to Switch B to the Master (CDHGFE). This circuitous path is non-optimal—and introduces latency and jitter.
- Failure of any node will force Spanning Tree to reconverge, and will increase network downtime after a failure.

Introduction to ESRP

ESRP combines the Layer 3 virtual router of VRRP with a Layer 2 virtual bridge. In ESRP, all links to the Slave are blocked. (In the diagram, links CD and GH will both be blocked.) If the Master node fails, the Slave in the diagram becomes the Master, and packets received over CD or GH will not be forwarded. This is done immediately, without waiting for a Layer 2 loop prevention protocol to reconverge. Furthermore, because ESRP only blocks forwarding from the Slave, it is guaranteed that traffic will follow the most efficient path to the Master.

Caveat

It is still technically possible to get a loop while running ESRP. ESRP prevents loops by having the Slave stop forwarding of all traffic. However, if Switch A and Switch B are connected (IJ), it is still possible to get a loop that does not require any forwarding via the Slave node. The loop here follows path IJFEBA.

Ensuring that Switch A and Switch B will never be connected together means that ESRP will prevent all loops—and convergence occurs much faster than Spanning Tree. (It is possible to run Spanning Tree on Switch A and Switch B in order to guarantee a loop-free topology—but doing so will increase convergence times and largely eliminate the benefits of using ESRP. Spanning Tree cannot be used on the Master or Slave, since ESRP and Spanning Tree cannot be configured to run on the same interface.)

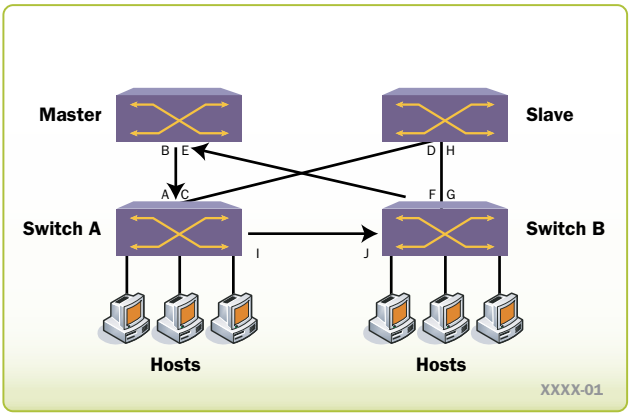

Figure 2

VRRP and ESRP Tracking

Tracking can be enabled on Extreme Networks switches, regardless of whether VRRP or ESRP is enabled. Tracking is used to check connectivity from the virtual router to the outside world. If connectivity is lost, the Master relinquishes control to the Slave—in the hopes that the Slave will still have connectivity. Tracking is an Extreme Networks extension to VRRP, so tracking can only be used when both the Master and the standby VRRP routers are Extreme Networks routers.

Three types of tracking are employed: ping tracking, route table tracking, and VLAN tracking. Ping tracking ensures that the Master has connectivity to its upstream router by periodically pinging the upstream router's interface. If multiple pings fail, the Master relinquishes control to the Slave. Likewise, route table tracking ensures the existence of routes in the routing table, and relinquishes control to the Slave if these routes don't exist. Alternatively, one can track based on whether any RIP, OSPF, or BGP entries have been installed in the route table—a simple test that ensures whether a routing peer has been discovered. Finally, VLAN tracking relinquishes control to the Slave if there are no active ports left in that VLAN.

ESRP Options

Domains

ESRP makes use of the concept of domains. Using domains, a single instance of ESRP provides redundant routing for up to 64 different VLANs. As a result, it is less processor intensive (and requires fewer advertisements be sent: one advertisement for the whole domain instead of one advertisement per VLAN). Because of support for Domains, ESRP scales to support a much larger number of VLANs than VRRP.

Host-Attach Mode

In ESRP's normal mode of operation, the standby forwards no traffic. This prevents loops from occurring.

In some cases, however, it may be desirable to connect hosts directly to ESRP switch, without having the concern that the switch will become a Slave and cease forwarding traffic. Turning on host-attach mode allows the Slave to forward traffic on the specified "host attach" ports. Recall that, by default, the Slave does not forward any traffic.

Aging Out Forwarding Database (FDB) Entries in VRRP/ESRP

When an ESRP or VRRP failover occurs, downstream switches will have learned MAC and IP addresses for upstream switches on the port connected to the Master, and must relearn these addresses on the port connected to the Slave (since the Slave becomes the Master after a failover). A flow of bidirectional traffic allows the FDB entry to be relearned on the new Master port—but if there is no bidirectional traffic, it could take up to 5 minutes for the entry to age out and for communication to be reestablished.

In an all-Extreme Networks network, the Master sends out Extreme Discover Protocol (EDP) messages announcing that it is the Master. When a failover occurs, downstream Extreme Networks switches listening to the EDP messages (i.e., "ESRP-aware" switches) immediately discover the new Master—and eliminate all FDB entries and ARP table entries associated with the port connected to the former Master (now the Slave).

If downstream switches are not Extreme Networks switches, another mechanism is required to inform the switches of the new Master. VRRP sends out unsolicited ARP replies, which allow downstream switches to learn the port to which the new Master is connected. In VRRP, this is a simple process—because VRRP is a Layer 3 only protocol.

Because ESRP works at Layer 2 as well as Layer 3, switches configured for ESRP use a different method for notifying downstream switches that a failover has occurred. An EDP message is sent out to tell downstream ESRP-aware switches that they must flush their FDBs and ARP tables. For non-ESRP aware switches, the original Master (which has become a Slave, after the failover) brings down the Ethernet ports connected to all downstream switches. When link is lost, the downstream switches will then flush all Forwarding Database entries for those ports. To configure this, use the following command on the Master and Slave devices:

```
config vlan <name> add ports
[<portlist> | all] restart
```

Conclusion

ESRP and VRRP both have similar overall goals: to provide fast failover in the event that a node or link fails. At Layer 3, both protocols allow multiple nodes to act as a single "virtual router"—so that failure of one node is completely transparent to any connected hosts.

ESRP has the additional benefit of acting as a Layer 2 loop prevention protocol for a redundant, partially meshed configuration—typically eliminating the need to use another loop prevention protocol (such as Spanning Tree, EAPS, or 802.1w). Furthermore, ESRP scales better because of its support for Domains, and supports additional tracking capabilities not supported in VRRP. Because of these benefits, it is recommended that ESRP be used as the protocol of choice in Extreme Networks-based networks. For multiple-vendor networks, a combination of VRRP and Spanning Tree can be used, but convergence times in the event of failure may be significantly slower.

The table below summarizes the differences between ESRP and VRRP.

ESRP	VRRP
Provides both Layer 2 and 3 redundancy	Provides Layer 3 redundancy only; Layer 2 loop prevention requires Spanning Tree
Defined by Extreme Networks	Defined by IETF in RFC 3768
Supports domains	Does not support domains
Up to 3000 VLAN's per ESRP instance	1 VLAN per VRRP instance
Failover based on master/slave communications failure or tracking of pings, route table entries, or VLANs	Failover based on master/slave communi-cations failure or priority change.Tracking func-tionality provided in Extreme Networks implementation only

Table 1

www.extremenetworks.com

Corporate and North America
Extreme Networks, Inc.
3585 Monroe Street
Santa Clara, CA 95051 USA
Phone +1 408 579 2800

Europe, Middle East, Africa and South America
Phone +31 30 800 5100

Asia Pacific
Phone +65 6836 5437

Japan
Phone +81 3 5842 4011