NETWORKWORLD

CORE NETWORKING

By Scott Hogg, Network World
JAN 2, 2013 6:59 PM PST

About

Scott Hogg is a co-founder of HexaBuild.io, an IPv6 consulting and training firm, and has over 25 years of cloud, networking and security experience.

INSIGHTS

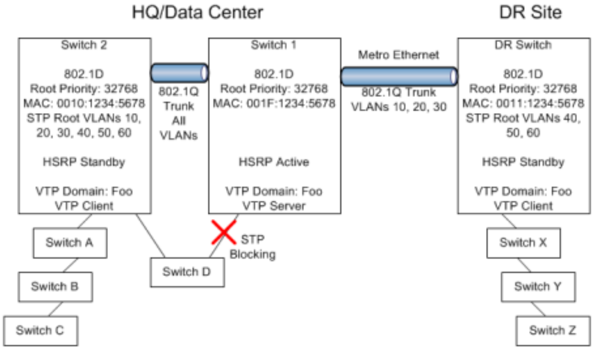
9 Common Spanning Tree Mistakes

Frequent spanning tree protocol misconfigurations cause network problems

Ethernet devices running the Spanning Tree Protocol (STP) have been implemented in networks since the early 1990s. Many organizations take STP for granted and do not configure it per industry best practices. STP errors are very common and during the past 15 years we have witnessed the same errors being made over-and-over again. For such a well established protocol, it is surprising that we have not progressed beyond these types of STP configuration mistakes. This article covers the most frequent STP errors and how to correct them.

The IEEE 802.1D Spanning Tree Protocol (STP) was invented by Radia Pearlman in 1985 when working at Digital Equipment Corporation (DEC). STP is a layer-2 protocol that runs between bridges to help create a loop-free network topology. Bridge Protocol Data Units (BPDUs) are packets sent between Ethernet switches (essentially multi-port bridges) to elect a root bridge, calculate the best path to the root and block any ports that create loops. The resulting tree, with the root at the top, spans all bridges in the LAN, hence the name: spanning tree. If you want to understand STP you should read the Algorhyme poem by Radia.

Spanning tree works efficiently at preventing loops with the default configuration settings. Thus, many people forget to adjust any parameters and simply accept the defaults. This leads many people to ignore STP in their network designs and, after many years, organizations are surprised to discover they have network issues related to spanning tree. There have been many optimizations to STP, but, if they have not been configured, the network is not benefiting from these new features.



This is picture of a typically misconfigured spanning tree environment that illustrates many of the common mistakes that are covered in this article.

1 - No Root Bridge Configured

Many organizations take spanning tree for granted and simply accept the default configuration settings. This leaves all switches in the environment using the default root bridge priority of 32768. If all switches have the same root bridge priority, the switch with the lowest MAC address will be elected as the root bridge. Many networks have not been configured with a single switch to have a lower root bridge priority which would force that core switch to be elected as the STP root for any or all VLANs. In this situation, it is possible that a small access-layer switch with a low MAC address could be the STP root. This situation would add some performance overhead and make for longer convergence times because of the root bridge reelection.

As seen in the above picture, the switch that is the STP root is actually core Switch 2 because it happens to have a lower MAC address than core Switch 1.

It is a best practice to configure the "main" (core) switches with lower STP priorities so that one will be the root bridge and any other core bridges will have a slightly higher value and take over should the primary core bridge fail. Having "tiered" STP priorities configured on the switches determines which switch should be root bridge in the event of a bridge failure. This makes the STP network behave in a more deterministic manner.

On the core Cisco switch you would configure the primary root switch with this command:

```
Core-Sw1(config)# spanning-tree vlan 1-4096 root primary
```

On the core Cisco switch you would configure the secondary root switch with this command:

```
Core-Sw2(config)# spanning-tree vlan 1-4096 root secondary
```

The net effect from these two commands will set the primary switch root bridge priority to 8192, and the secondary switch root bridge priority to 16384.

2 - Use of IEEE 802.1D and not Rapid-STP

The classic IEEE 802.1D protocol has the following default timers: 15 seconds for listening, 15 seconds for learning, 20 second max-age timeout. All switches in the spanning tree should agree on these timers and you are discouraged from modifying these timers. These older timers may have been adequate for networks 10 to 20 years ago, but today this 30 to 50 seconds of convergence time is far too slow.

Today, many switches are capable of Rapid Spanning Tree Protocol (IEEE 802.1w), but few network administrators have enabled it. RSTP vastly improves convergence times by using port roles, using a method of sending messages between bridges on designated ports, calculating alternate paths, and using faster timers. Therefore, organizations should use RSTP when they can. If your organization still has switches that cannot run RSTP, don't worry, the RSTP switches will fall back to traditional 802.1D operation for those interfaces that lead to legacy STP switches.

3 - Blocked Uplinks

Spanning Tree's job is to prevent loops from forming. It does this by learning about sub-optimal paths to the root and placing these less desirable links into blocking mode. If there are multiple parallel paths between switches, then one of them would be selected to be in blocking mode to prevent a loop between the two switches. This would make having multiple uplinks only good for failover for the primary link and not provide increasing bandwidth along that path.

In the picture above you can see this situation for Switch D. Its link toward Switch 2 is the optimal path to the root bridge but its suboptimal path to Switch 1 is in STP blocking state. Therefore, only one link's bandwidth is available for upstream communications.

I we wanted to be able to utilize both uplinks to forward traffic and increase our bandwidth, then we could use some form of link aggregation like a port-channel/EtherChannel (LACP (IEEE 802.3ad), PAgP) or some form of multi-chassis port-channel (MC-LAG IEEE 802.3AX/AY) or use Cisco Nexus switches with a virtual Port Channel (vPC).

Another option would be to use stackable switches and configure each uplink port to connect to a different switch in the stack. Since the stack would be configured as if it was one switch, a port-channel could be used and both links would be treated like a single link from a spanning tree perspective and both links would be available for forwarding traffic. Other options such as Cisco's 6500 Virtual Switching System (VSS) would be able to accomplish this same result.

4 - Exceeding STP Maximum Dimensions

We have all been in cities that lacked coordinated civic planning and their roads are always congested and unnavigable. Similarly, networks often grow organically like aspen trees. Networks get new devices added to them, but they are seldom re-architected unless a completely new network is purchased. As new switches are added to a LAN environment the spanning tree evolves over time.

Large networking environments supporting applications that rely on layer-2 connectivity across the entire network should be aware of this growth. These organizations can experience problems if their topology exceeds STP's maximum dimensions. The 802.1D specifications recommends that a spanning tree have no more than seven bridge hops. This can easily occur when there are many "daisy-chained" switches. See the picture above and see how even a simple topology can exceed this maximum spanning tree dimension.

Organizations like hospital networks or college campuses with large expansive LAN environments that extend a single VLAN across the entire breadth of the network should be cognizant of their spanning tree dimensions. These organizations should periodically check their spanning tree dimensions by documenting their LAN switched environments and by looking for excessively frequent Topology Change Notifications (TCNs).

5 - VTP Domains

Organizations often have difficulties related to VLAN Trunking Protocol (VTP) and how it relates to STP. VTP is a mechanism to aid in the creation and maintenance of VLANs in a single LAN switched environment. The VTP server can create new VLANs and the VTP clients will then be automatically configured with those new VLANs. Then ports on those VTP client switches can be assigned to this new VLAN. VTP helps keep VLAN numbering consistent in a LAN switched environment. There have been a long history of issues related to VTP and many recommend configuring VTP in transparent mode.

Some organizations take the time to configure their VTP domain and configure a VTP server and clients. Some organizations also use the same VTP domain name across all switches, even at all their locations. This can start to cause problems when layer-2 Metro-Ethernet services are used and configured as an 802.1Q trunk.

This problem is illustrated in the diagram above. The switches in this example all have the same VTP domain name. The connection between the data center and the DR site uses an 802.1Q trunk but only three VLANs are allowed on this trunk. Therefore, the switch at the DR site knows about all the VLANs, and uses spanning tree to determine that the Metro-Ethernet link is the path toward the root for the VLANs 10, 20 and 30. However, the DR switch believes that it is the STP root for the other VLANs that are not used on the trunk. This could potentially cause a problems because now the environment has two switches that believe they are the STP root for VLANs 40, 50, 60.

Organizations that use VTP should use it carefully and know which switches are VTP servers or clients, use a VTP password, delete configuration and VTP information from switches removed from service, and consider disabling VTP where it is not needed.

6 - STP Incongruence with HSRP

Many organizations have redundant core switches that are also the layer-3 default gateway for computers on the connected LANs. First Hop Redundancy Protocols like HSRP, VRRP, GLBP, among others, provide default gateway redundancy for hosts that are configured with only a single default gateway IP address.

The issue arises when the HSRP active default gateway is not the same Layer2/3 switch that is root of the STP for that VLAN. The diagram above shows that Switch 1 is the HSRP active router, but it is not the STP root for any VLAN. Switch 2 is the STP root, but it is configured as the HSRP standby router. This creates non-optimal traffic paths which can lead to higher congestion on the inter-core-switch trunk.

Organizations that use a First Hop Redundancy Protocol should make sure that there is alignment between the active default gateway and the STP root.

7 - Failure to Control STP

Because so many organizations simply accept the switch's manufacturer default settings for spanning tree they are not optimally controlling STP. Organizations may not be configuring spanning tree to prevent against an inadvertently added rogue switch from creating a loop. Many organizations use Cisco's PortFast interface settings to help bring up switchports quickly for ports connected to computers that we know do not run STP. It does not make sense to have the port connecting to a computer waiting through the listening and learning states before activating the interface.

It is a best practice to use PortFast with BPDU-Guard so that it defensively shuts down the port if a BPDU is received on that interface.

The Cisco IOS global command to active this feature is:

```
Core-Sw1(config)# spanning-tree portfast edge bpduguard
```

The Cisco IOS interface configuration command to active this is:

```
Core-Sw1(config-if)# spanning-tree bpduguard enable
```

If a switch has any port-channels configured, then it is a good idea to configure EtherChannel guard.

The Cisco IOS global command to active this feature is:

```
Core-Sw1(config)# spanning-tree etherchannel guard misconfig
```

Organizations should also use Root Guard on all access-switch ports connecting to servers.

The Cisco IOS interface configuration command to active this is:

```
Core-Sw1(config-if)# spanning-tree guard root
```

Sometimes blade servers have embedded Ethernet switches and these should also be factored into the STP design and configuration. The configuration of these switches should be treated the same as any other STP device and its configuration should complement the other switches in the environment.

8 - Inconsistent Spanning Tree Metrics

Traditionally, spanning tree has used a 16-bit value for the link cost used by bridges for calculating the shortest path to the root. With these older 16-bit metrics, a 10Mbps link would have a cost of 100 and a 1Gbps link would have a cost of 4. However, link speeds have outgrown these metrics and there are now a 32-bit long path cost. With the newer 32-bit metrics, a 1Gbps link would have a cost of 20,000 a 10Gbps link would have a cost of 2,000 and a 100Gbps link would have a cost of 200.

To enable the long path cost on a Cisco switch, simply enter this global configuration command.

Page 1 of 2 ›

Verizon, Amazon team to offer 5G edge cloud
Homepage

How cloud providers' performance differs
Homepage

Amazon joins the quantum computing crowd with
Homepage