Comparison between PVST+ and MSTP

A Dell Technical White Paper

Dell Engineering
November 2014

A Dell Best Practices

Revisions

Revision	Date	Description	Author/Editor
4	11/05/2014	Comments incorporated.	Chris Patrick, Patti Stone, Sreepad Putti Subba, Mike Matthews
3	08/18/2014	Comments incorporated.	Shuaib Ilyas
2	07/20/2014	Comments incorporated.	Amol Rawal
1	06/16/2014	Initial release	Shuaib Ilyas

Table of contents

1 Introduction

The Spanning Tree Protocol (STP) is used to protect networks from loops. Over time, there have been several enhancements to this protocol, including Per VLAN Spanning Tree Plus (PVST+) and Multiple Spanning Tree (MST). PVST+ is a Cisco proprietary protocol while MSTP is based on the IEEE standard 802.1s. In this white paper, the basic workings of STP are discussed and a comparison between PVST+ and MSTP is provided.

2 Overview of STP

In order to have reliable networks, network switches are equipped with redundant equipment including power supplies, CPUs, inter switch links etc. Whenever there are redundant links between the Layer 2 devices in a network, they cause traffic loops that can bring down the network. Spanning tree is a layer two protocol designed to protect a network from Ethernet network loops. After convergence (all devices agree on what the network topology looks like), the whole network becomes a logical tree, as the redundant paths are disabled until required (due to a link or switch failure in the network). From any source to destination, there is only one logical path in the network. The protocol forces the redundant ports to block all the traffic. In this way, even if there are redundant physical paths in the network, there are no traffic loops in the network. It is only when the active port goes down that the protocol reconverges and the blocking port starts forwarding the traffic in the new logical topology.

The issue of traffic looping exists because Ethernet does not have a header equivalent to TTL (Time to Live) which exists in the Internet Protocol (IP) world (Layer 3 in the OSI model). In the case of IP, if there is a loop in the network, it will eventually die. When a router forwards a packet, it decrements the TTL value in the IP header by 1. So, if a router receives a packet with a TTL value of zero, it will drop the packet and there will not be any loops in the network. In case of Ethernet, there is no such mechanism. So, if a loop starts in the network, it will continue forever and cause the CPUs to spike to 100% unless the physical link is disconnected. In addition, since Ethernet is not reliable by design, it will cause traffic to be dropped because of traffic congestion.

In case of Ethernet, if a frame with an unknown destination is received by a switch, it floods the frame to all its ports except the incoming port.

Figure 1 provides a reference for the following explanation of how a loop is created in an Ethernet network. If the Client needs to communicate to the Server, it will send an Ethernet frame to Switch 4 using its MAC address as the source and the broadcast address (FF:FF:FF:FF:FF:FF) as the destination MAC address. Switch 4 will learn the MAC address of the client and since it does not know the MAC address of the server, it will flood it, i.e. send it to all the ports except the port on which the frame was received. So, this frame will be sent to both Switch 1 and Switch 3. Next, Switch 1 will flood the frame to both Switch 2 and Switch 3. Likewise, the frame will be flooded from Switch 3 to both Switch 1 and Switch 2. The frame received by Switch 1 will then be flooded again to Switch 2 and Switch 4. Notice that Switch 4 (the switch that initiated the frame) has received the same frame and the frame will start looping in the network.

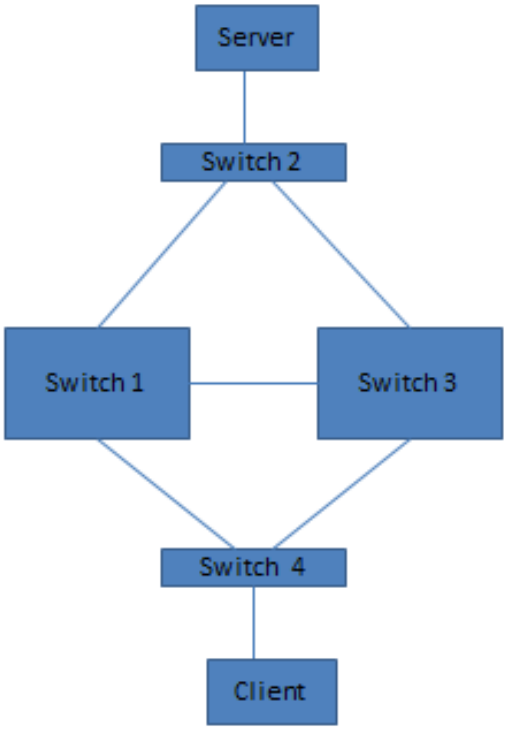

Figure 1 How a Loop is created in an Ethernet Network

3 Basic Spanning Tree Process

The spanning tree protocol uses BPDUs (Bridge Protocol Data Units) to create a loop free topology. BPDUs are control packets which help to calculate the best path in a network and disable secondary paths resulting in a loop free topology.

Spanning tree uses the following steps to create a loop free topology:

1. Root switch/bridge election.
2. Root port election.
3. Designated port election.
4. Blocking all non-edge ports.

3.1 Root Switch/Bridge Election

The switch with the lowest bridge ID becomes the root bridge in the network. The bridge ID consists of the bridge priority and the MAC address of the switch. The default bridge priority is 32768, so if bridge priority is kept as default on all the switches in the network, the switch with the lowest MAC address will become the root switch.

Selecting the appropriate root bridge is extremely important. Because switches manufactured earlier have lower MAC addresses compared to newer switches, if the default bridge priority is not changed an older switch that is not capable of handling all the traffic and not in the optimum traffic path will be elected as root bridge. If a newer switch with more capabilities (i.e. CPU power, memory and/or location in the network) needs to be set up as the root bridge, its bridge priority should be lowered so the election is based on the lowest priority value instead of the MAC address.

3.2 Root Port Election

The root port is the upstream port towards the root bridge. The election of the root port is based on the following criteria:

1. The lowest root path cost to get to the root bridge.
2. In the event of a tie, the lowest upstream Bridge ID is used.
3. In the event of a tie, the lowest upstream port ID is used.

3.3 Designated Port Election

The designated port is the downstream port away from the root bridge. The election of the designated port is based on the following criteria:

1. The lowest root path cost to get to the root bridge.
2. In the event of a tie, the lowest upstream bridge ID is used.
3. In the event of a tie, the lowest upstream port ID is used.

3.4 Blocking all Non-Edge Ports.

Once the root bridge, root ports and designated ports are elected, the rest of the ports are put in blocking state for all traffic.

3.5 Spanning Tree Example

In the beginning, every switch claims to be the root and sends BPDU to all the other switches connected to it. Any switch receiving the BPDU compares its bridge id to the one it received. The bridge id is made of 2 parameters; Bridge priority and MAC address. The lowest bridge id switch wins the root bridge election. If a switch receives a BPDU with lower bridge id, it stops sending its own inferior BPDU. In this way, once the topology converges, only the root bridge sends the BPDUs and the rest of the switches forward the BPDU received from the root bridge. By this mechanism, all switches in the network agree on one Root Bridge.

Consider the following topology (Figure 2):

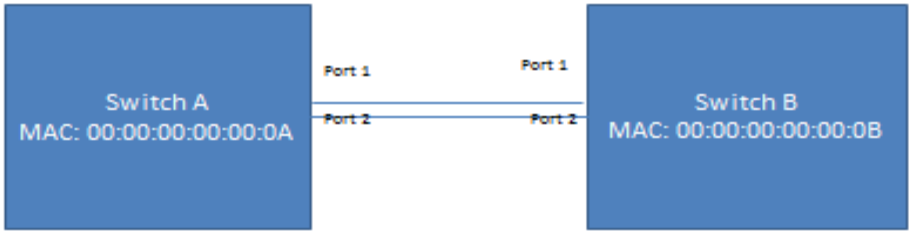

Figure 2 Root Bridge Election

Since the default bridge priorities are being used, the switch with the lower MAC address wins the election of the root bridge. In this topology, Switch A becomes the root switch of the network since it has the lower MAC address.

Next, all the non-root switches determine the root port that will be used to reach the root switch. The root port is the port on each switch that receives the best BPDU i.e. lowest path cost to the root bridge. In this case, Switch B has two links to Switch A, so one of the ports will become the root port. Assuming the two links have the same bandwidth (lowest path cost to the root bridge), the tie breaker will be lowest upstream port ID. So, port 1 on Switch B becomes the root port.

Next, the designated ports are selected in each segment. In our example, since Switch A is root bridge, all the ports on Switch A will become designated ports (lowest path cost to the root bridge) and go into the forwarding state.

Finally, port 2 on Switch B will go into blocking state as it is neither root nor designated port. So spanning tree will make the logical topology of the network as shown in Figure 3.

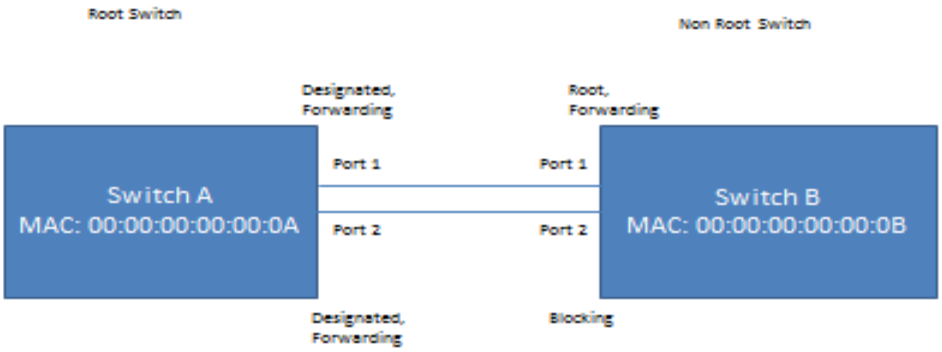

Figure 3 Port States after Convergence

4 Overview of PVST+

As the name suggests, in Per VLAN Spanning Tree Plus (PVST+), a separate instance of spanning tree runs for every VLAN. So, if there are 10 VLANs created on a switch, there will be 10 instances of spanning tree running independent of each other. Since each VLAN has its own instance of spanning tree, the root switch can be different for each VLAN and accordingly, the traffic forwarding path as well. PVST+ uses 802.1Q headers in Ethernet frames. It sends untagged BPDUs using the IEEE multicast MAC address 01:80:C2:00:00:00 for interoperability with STP, and It sends PVST+ BPDUs using Cisco's multicast MAC address 01:00:0C:CC:CC:CD.

4.1 PVST+ Best Practice

Typical best practice for PVST+ is to make one switch root for odd number of VLANs and the other switch root for even number of VLANs. This way, both the links will be used at the same time and traffic will be load balanced. This assumes the same amount of traffic is occurring on all the vlans.

4.2 PVST+ Example

The sample configuration below is for power connect switches.

In the whole network, only two switches require configuration, the root switch of the even VLANs and the root switch of the odd VLANs (Figure 4). So, it is much simpler configuration compared to MSTP.

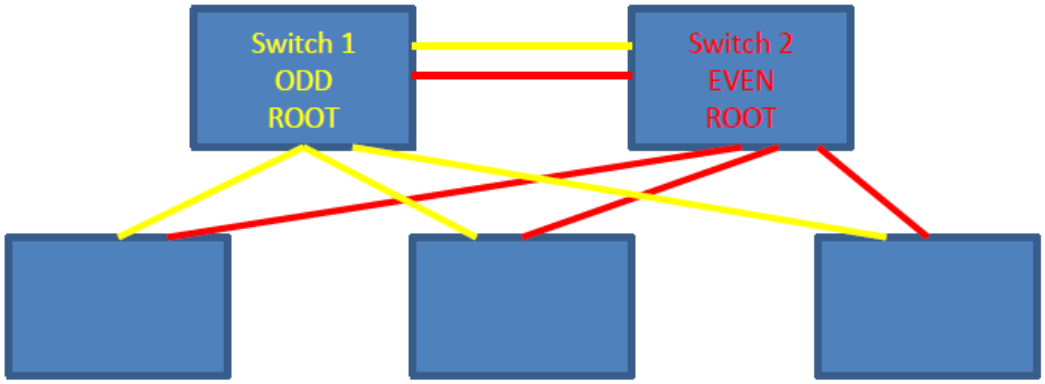

Figure 4 PVST+ Topology

The commands necessary to configure PVST+ on switch 1 and 2 are shown in Table 1.

Switch 1 (Root for Odd VLANs)	Switch 2 (Root for Even VLANs)	Description of Commands
`spanning-tree mode pvst`	`spanning-tree mode pvst`	Enabling pvst mode of spanning tree on the switch.
`spanning-tree vlan 1,3,5,7,9 priority 8192`	`spanning-tree vlan 1,3,5,7,9 priority 16384.`	Configuring bridge priority for odd VLANs. Switch 1 becomes root bridge for odd VLANs because of the lower priority configured.
`spanning-tree vlan 2,4,6,8 priority 16384`	`spanning-tree vlan 2,4,6,8 priority 8192`	Configuring bridge priority for even VLANs. Switch 2 becomes root bridge for even VLANs because of the lower priority configured.

Table 1 Command list for PVST

5 Overview of MSTP

One disadvantage of PVST+ is the instances of spanning tree are equal to the number of vlans. Since every instance of spanning tree is CPU intensive, PVST+ requires a lot of resources to run on a network with a large number of vlans.

In the case of Multiple Spanning Tree Protocol (MSTP), there are multiple instances of spanning tree but not every vlan has to be in a different instance of spanning tree. In a typical example, if a physical topology has two variations, only two instances of spanning tree are needed. Half of the vlans can be grouped in the first instance of spanning tree and the other half in the second instance. This way, the switch only needs enough resources to run two instances of spanning tree. Furthermore, MSTP is an IEEE standard so different vendor's products can interoperate. In case of PVST+, it is Cisco proprietary.

5.1 MSTP Best Practice

Typical best practice for MSTP is to use two instances of spanning tree, one for the odd vlans and the other for the even vlans.

5.2 MSTP Example

The sample configuration shown in Figure 5 is for Powerconnect switches.

Every single switch in the network has to be configured for the following parameters:

1. Region name
2. Rev #
3. VLAN to instance mapping.

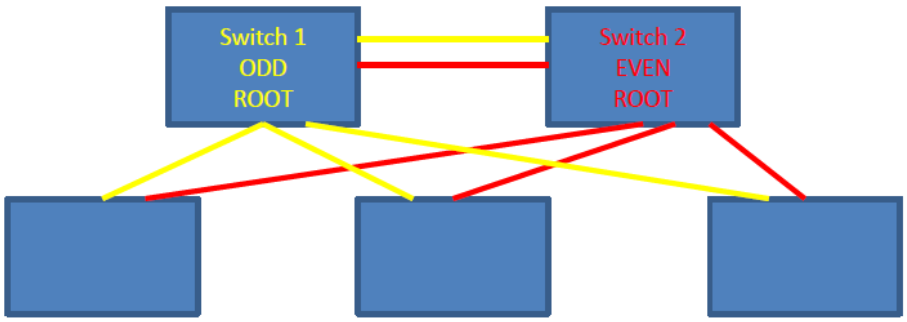

Figure 5 MSTP Topology

The commands necessary to configure MSTP are shown in Table 2.

Switch 1 (Root Yellow)	Switch 2 (Root Red)	All non-root switches	Description of commands
`spanning-tree mode mst`	`spanning-tree mode mst`	`spanning-tree mode mst`	Configuring spanning tree mode to MSTP on the switch.
`spanning-tree mst conf`	`spanning-tree mst conf`	`spanning-tree mst conf`	MSTP configuration mode to configure MSTP parameters.
`revision 1`	`revision 1`	`revision 1`	Configuring revision # which is a required MSTP parameter.
`name YellowRed`	`name YellowRed`	`name YellowRed`	Configuring Region name of the MSTP domain.
`instance 1 add vlan 1,3,5,7,9`	`instance 1 add vlan 1,3,5,7,9`	`instance 1 add vlan 1,3,5,7,9`	Associating VLANs 1,3,5,7,9 to spanning tree instance 1.
`spanning-tree mst priority 8192`	`spanning-tree mst priority 16384`		Configuring bridge priority.
`spanning-tree mode mst`	`spanning-tree mode mst`		Configuring spanning tree mode to MSTP on the switch.
`spanning-tree mst conf`	`spanning-tree mst conf`		MSTP configuration mode to configure MSTP parameters.
`instance 2 add vlan 2,4,6,8`	`instance 2 add vlan 2,4,6,8`	`instance 2 add vlan 2,4,6,8`	Associating VLANs 2,4,6,8 to spanning tree instance 2.
`spanning-tree mst priority 16384`	`spanning-tree mst priority 8192`		Configuring bridge priority.

Table 2 Command list for MSTP

6 Comparison between PVST+ and MSTP

6.1 Proprietary vs Open Standard

PVST+ is a Cisco proprietary protocol, while MSTP is an open standard protocol based on the IEEE standard 802.1s. So, in multi-vendor environments, MSTP is the preferred option because of interoperability.

6.2 Number of Spanning Tree Instances

In the case of PVST+, there are as many instances of spanning tree as the number of VLANs. This can result in PVST+ being very resource intensive. In the case of MSTP, theoretically, the number of spanning tree instances can be the same as the number of VLANs, but for all practical purposes, the number of spanning tree instances are restricted to number of physical topologies which are significantly less than the number of VLANs in the network.

6.3 Configuration Simplicity

Configuration is simpler in case of PVST+. It is required only on the root switches in the network. In the case of MSTP, configuration is relatively complicated and is required on every single switch in the network.