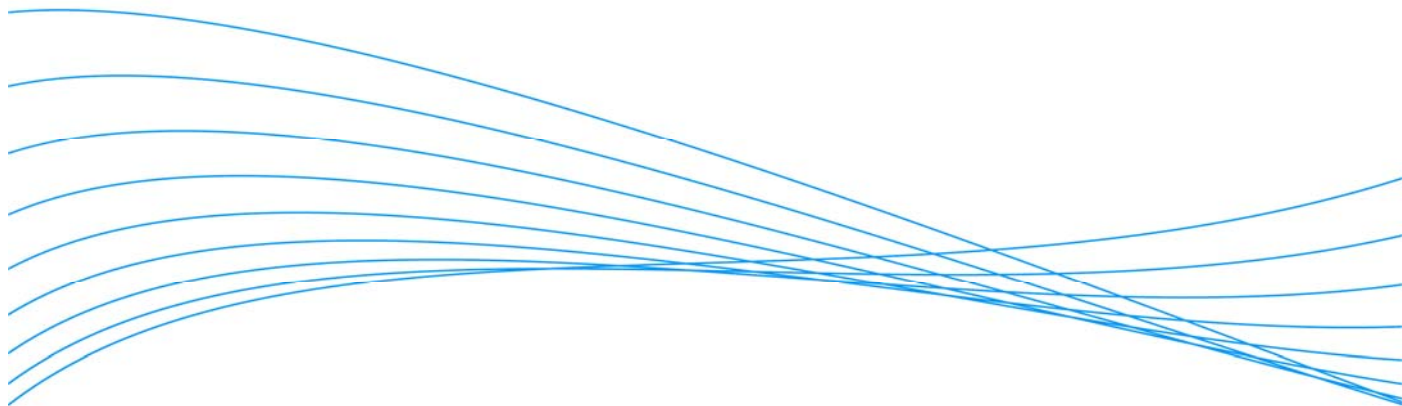


Interoperating with Cisco's RPVST+ Spanning Tree Protocol in a High Availability Topology



Interoperating with Cisco RPVST+	2
The Problem:	2
Solution 1 – Go with open standards!	3
Description	3
Positives	4
Negatives	4
Solution 2 – Let RPVST+ do the work	4
Description	4
Positives	4
Negatives	4
Solution 3 – Unified CST approach	5
Description	5
Positives	5
Negatives	5
Recommendations	5
Terminology	6

Interoperating with Cisco RPVST+

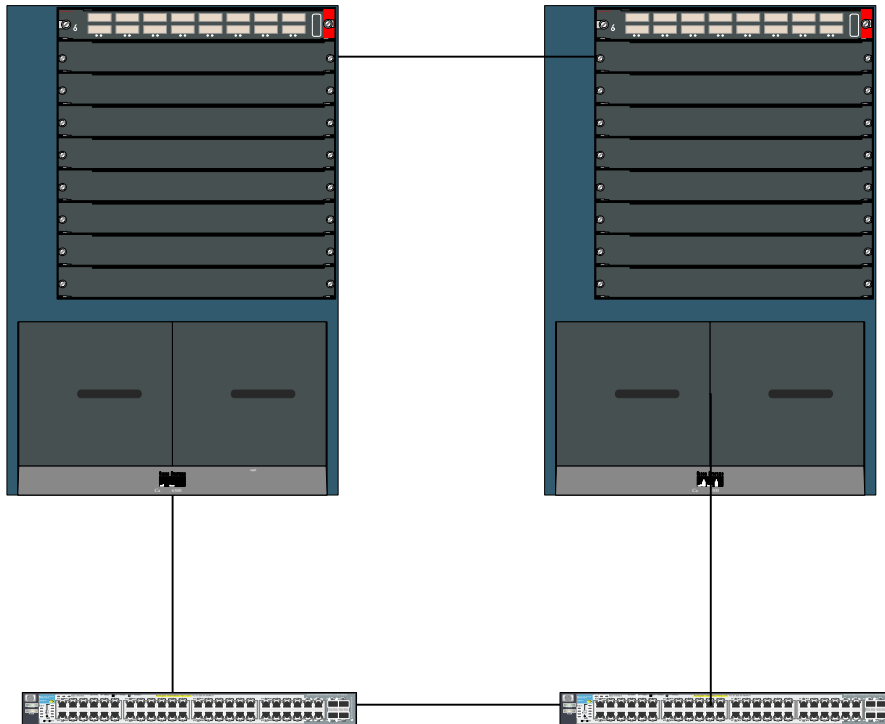
Cisco Proprietary Rapid Per-VLAN Spanning Tree (RPVST+) protocol can create problems for networking equipment, such as ProCurve devices, that use standards-based solutions. This document will address some of the problems created by RPVST+, and suggest ways to mitigate the issues of slow convergence times and spanning tree loops.

One common scenario where this topology might occur is when customers deploy two access switches either into a cabinet or nearby server to provide Layer 2 redundancy in case of access switch or NIC failure. In this high-availability environment failover and failback times are critical, as they can cause vital systems to fail even with very short reconvergence times.

The Problem:

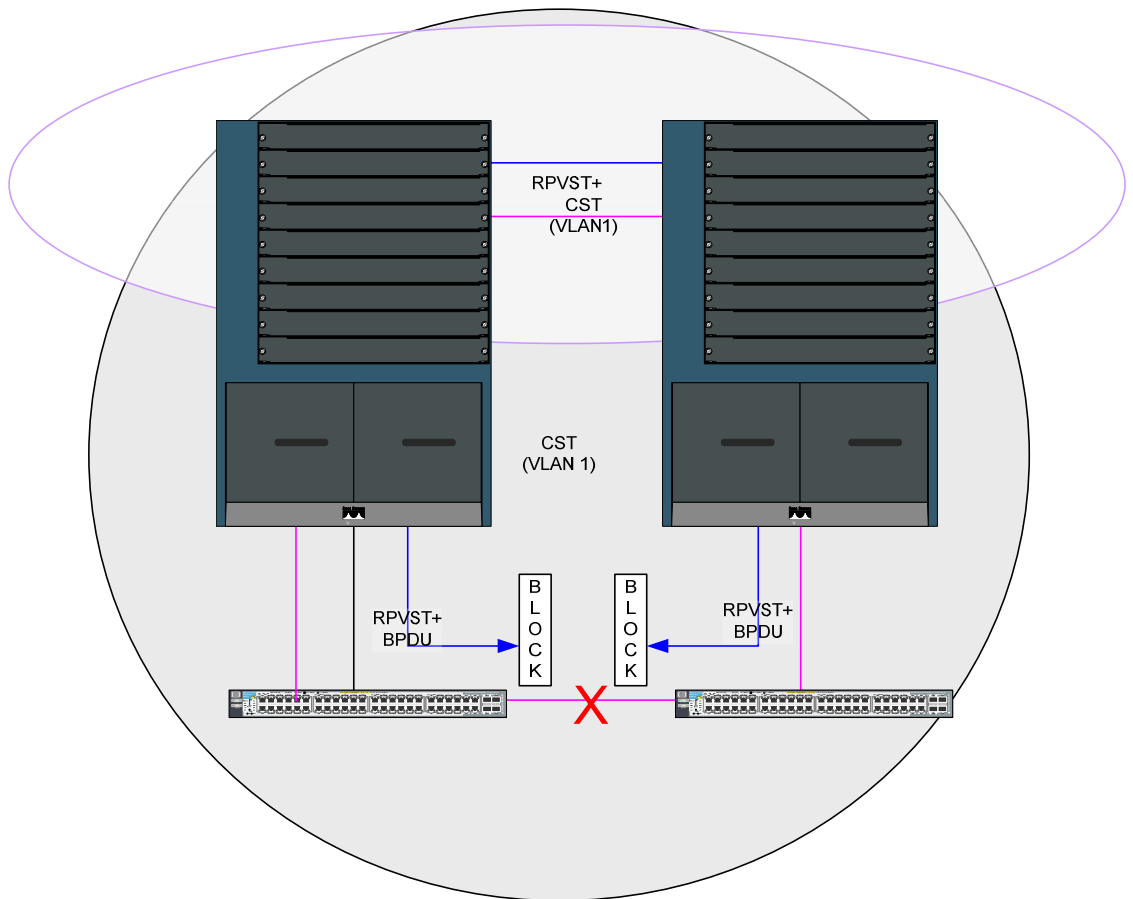
You are using Cisco devices in the core of your network with RPVST+ and a standards-based spanning tree switch on the edge. In our example here is a ProCurve 3500zl using 802.1s (MSTP), which is the default. 802.1s, when it hasn't been configured, operates exactly the same as 802.1w (RSTP), so this solution applies to both RSTP and MSTP.

With RPVST+, the Cisco spanning tree has a different topology for each VLAN, but VLAN 1 is given special treatment by being designated as the CST¹ (the instance at which RPVST+ is interoperable with RSTP/MSTP).



As this diagram demonstrates, there is a potential loop. Using normal VLAN tagging (trunking on Cisco) and default spanning tree commands at all places – but specifying the root as a Cisco – we get a network that operates and is loop-free (the spanning tree is blocking, therefore preventing the loop). It is interesting to note that the mechanism is actually doing the traffic blocking. In this situation, there are two switches running RPVST+ on all VLANs, and in this scenario, there is only one link between them. Looking at the Cisco devices, all ports are forwarding. From the RPVST+ point of view, there are no loops in the network except for VLAN 1. The following diagram illustrates why this is the case.

¹ This is always VLAN 1, therefore VLAN 1 should always be the VLAN used when connecting to non-Cisco devices.



The link between the ProCurve 3500zIs is blocking because, according to CST rules, that is the most appropriate place for them to do so. This in turn is blocking RPVST+ BPDU from being forwarded across the ProCurve Switches². This is the designed behavior of spanning tree. This doesn't prevent interoperability between Cisco RPVST+ and standards-based STPs, although it does pose some difficulties.

One thing to beware of is that for this to operate correctly, according to Cisco, the CST Root should always be within the RPVST domain. This is because CST blocking will not affect any VLAN other than VLAN 1, and therefore, while the CST will be trying to block VLAN 10, RPVST+ will be forwarding for VLAN 10. This is a limitation of RPVST+.

Given the diagrams, this will work fine during normal operations. And if a failure occurs, that will be picked up as well, and traffic should flow fairly quickly (constrained by the speed at which rapid spanning trees can converge). However, there is a problem that occurs when you recover from a failure. This is caused by RPVST+ being thrown into a full spanning tree state machine cycle because it is not receiving BPDUs on non-CST VLANs on ports connecting to the ProCurve 3500s. Since the RPVST+ VLANs are not receiving BPDUs from the ProCurve devices, it will not go into forwarding until its forward-delay timer has expired.

There are three solutions to this situation:

Solution 1 – Go with open standards!

Description

The first and most reliable solution by far is to migrate from a RPVST+ solution to a complete MSTP or RSTP standards-based solution. This will allow for the most reliable and integrated solution as well as the most deterministic solution. Troubleshooting is easier, and with MSTP, the same results as RPVST+ can be achieved. However, this isn't always an option due to IOS constraints or policy choices by network administrators.

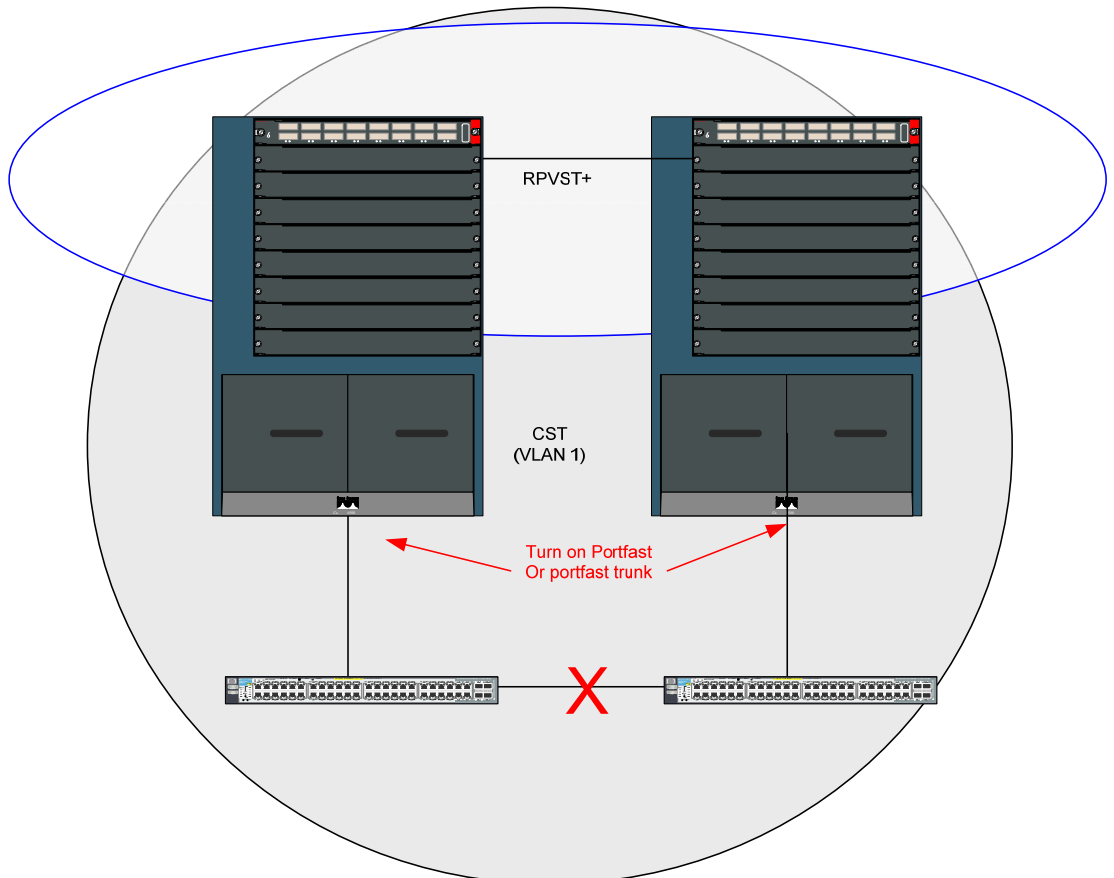
² ProCurve devices see these are Multicast packets, and forward them as they would or wouldn't any other packet

Solution 3 – Unified CST approach

Description

Another solution is to simply short circuit the state machine process on the RPVST+ VLANs by enabling the “spanning-tree portfast” command on the links to the ProCurve 3500s. If the Cisco device is configured for 802.1q trunking, the “spanning-tree portfast trunk” command can be used. This allows a port to go into a forwarding state immediately while remaining able to react to network loops if it receives a BPDU.

This creates a blocking in the most optimal place, which is between the two 3500s. This is most optimal because it is the furthest from the core, and also, because on a failure of an edge switch, it has the least impact for hosts more than likely trying to get to the distribution/core layer.



Positives

This is the optimal spanning tree blocking point because on an edge switch failure, the forwarding of the other switch is not impacted; plus, no ports have to change states in the data path. The only solution this doesn't necessarily help is when the link is lost between the two core switches. That is not something an edge switch design can accommodate. To minimize the number of large impacts to the topology, core boxes with higher availability are recommended, as is redundant power.

Negatives

There are concerns about portfast being enabled, as that may cause a very brief network loop when a link arises. However, because of the rapid convergence times of 802.1w (and RPVST+), this will be blocked quickly, if necessary. The approach is treating the whole system like a CST, but the RPVST+ choices can be made as well.

Note: Receiving a BPDU on a portfast-enabled port will cause PVST to wait 2 x forward-delay before transitioning to the forwarding state. Therefore, it is recommended that the ProCurve switches be configured to filter the multicast PVST MAC address (01:00:0c:cc:cc:cd).

Recommendations

All these approaches depend on a common design. Large spanning tree topologies should be planned and blocking points determined in advance. The simpler the solution, the more reliable and scalable it will be.

For most cases where high availability is a must, and optimal switching also is desired, Solution 1 is the best bet. However, Solution 3 can accommodate the needs of high availability and robustness, as well. Meanwhile, Solution 2 has several potential pitfalls that make it a little bit riskier and less optimal.

Lastly, a note about filtering RPVST+ BPDUs. While many of the ProCurve products can filter out RPVST+ BPDUs, the main benefactor of this is RPVST+ devices, as their logs will fill up with messages about differing bridge ID's being received on a single port. You should contact Cisco Technical support for a recommendation on this. It should only be done if seeing multiple Bridge IDs coming in on a single port is causing a problem, which would be a Cisco specific issue.

Terminology

STP	Spanning Tree Protocol
CST	Common Spanning Tree, usually 802.1w or 802.1d compliant
PVST	Per-VLAN Spanning Tree (Cisco Proprietary)
RPVST	Rapid Per-VLAN Spanning Tree (Cisco Proprietary)
(R)PVST+	PVST with CST interoperability
802.1w	Standards-based rapid spanning tree protocol
802.1s	Standards-based multi-instance spanning tree protocol
Spanning Tree	Layer 2 loop detection and avoidance mechanism
Blocking	A spanning tree state that prevents a network loop
Portfast	A Cisco feature that makes a port bypass the state machine process
Trunking	In Cisco devices, this means a 802.1q tagged link
Trunking (2)	In ProCurve devices, this refers to port aggregation
BPDUs	Bridge Protocol Data Unit, the method of STP communication
802.1q	Standards-based method of putting multiple VLANs on a link
Tag	The method of implementing 802.1q

To find out more about
ProCurve Networking
products and solutions,
visit our web site at

www.procurve.com



© 2007 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice. The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.

4AA1-5427ENN, 10/2007