

Cisco – Understanding Multiple Spanning-Tree Protocol (802.

Table of Contents

<u>Understanding Multiple Spanning–Tree Protocol (802.1s)</u>	<u>1</u>
<u>Introduction</u>	<u>1</u>
<u>PVST+ Case</u>	<u>2</u>
<u>Standard 802.1q Case</u>	<u>2</u>
<u>MST Case</u>	<u>2</u>
<u>MST Region</u>	<u>3</u>
<u>MST Configuration and MST Region</u>	<u>3</u>
<u>Region Boundary</u>	<u>4</u>
<u>MST Instances</u>	<u>4</u>
<u>IST Instances</u>	<u>5</u>
<u>MSTIs</u>	<u>5</u>
<u>Common Misconfigurations</u>	<u>6</u>
<u>IST Instance is Active on All Ports, Whether Trunk or Access</u>	<u>6</u>
<u>Two VLANs Mapped to the Same Instance Will Block the Same Ports</u>	<u>7</u>
<u>Interaction Between the MST Region and the Outside World</u>	<u>8</u>
<u>Recommended Configuration</u>	<u>9</u>
<u>Alternate Configuration (Not Recommended)</u>	<u>10</u>
<u>Invalid Configuration</u>	<u>11</u>
<u>Migration Strategy</u>	<u>11</u>
<u>Conclusion</u>	<u>12</u>
<u>Tools Information</u>	<u>12</u>
<u>Related Information</u>	<u>12</u>
<u>Related Topics</u>	<u>12</u>
<u>Additional Documentation</u>	<u>12</u>

Understanding Multiple Spanning–Tree Protocol (802.1s)

Introduction

PVST+ Case

Standard 802.1q Case

MST Case

MST Region

MST Configuration and MST Region

Region Boundary

MST Instances

IST Instances

Common Misconfigurations

IST Instance is Active on All Ports, Whether Trunk or Access

Interaction Between the MST Region and the Outside World

Recommended Configuration

Alternate Configuration (Not Recommended)

Invalid Configuration

Migration Strategy

Conclusion

Tools Information

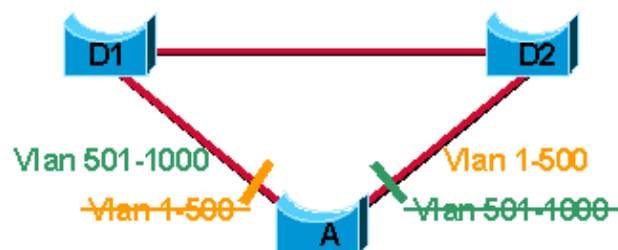
Related Information

Introduction

Multiple Spanning–Tree (MST) is a new Institute of Electrical and Electronics Engineers (IEEE) standard inspired from Cisco's proprietary Multiple Instances Spanning–Tree Protocol (MISTP) implementation. This document assumes that the reader is familiar with Rapid Spanning–Tree Protocol (RSTP) (802.1w), as MST heavily relies on this other IEEE standard. A pre–standard MST is implemented as of CatOS 7.1 and Native Cisco IOS® Version 12.1(11)EX. For more information on RSTP (802.1w), refer to the following document:

- Understanding Rapid Spanning–Tree Protocol (802.1w)

The following diagram shows a very common design featuring access Switch A with 1000 VLANs redundantly connected to two distribution Switches D1 and D2. In this setup, users connect to Switch A, and the network administrator typically seeks to achieve load balancing on the access switch uplinks based on even or odd VLANs, or any other scheme deemed appropriate.



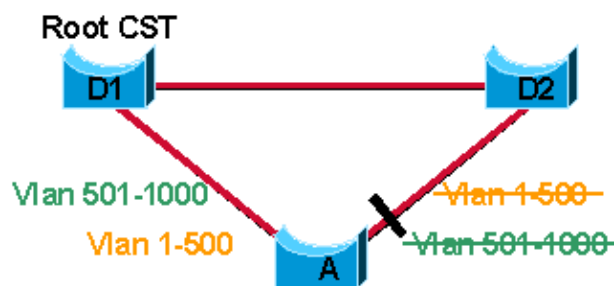
PVST+ Case

In a Cisco Per-VLAN Spanning-Tree (PVST+) environment, the spanning-tree parameters are tuned so that half of the VLANs are forwarding on each uplink trunk. This is easily achieved by electing Bridge D1 to be the root for VLAN 501-1000, and Bridge D2 to be the root for VLAN 1-500. The following are true for this configuration:

- In this case, optimum load balancing results.
- One spanning-tree instance for each VLAN is maintained, which means 1,000 instances for only two different final logical topologies. This considerably wastes CPU cycles for all the switches in the network in (addition to the bandwidth used by each instance sending its own Bridge Protocol Data Units (BPDUs)).

Standard 802.1q Case

The original IEEE 802.1q standard defines much more than simply trunking. It defines a Common Spanning-Tree (CST) that only assumes one spanning-tree instance for the entire bridged network, regardless of the number of VLANs. If the CST is applied to the topology of the above diagram, the result resembles the following diagram:



In a network running the CST, the following are true:

- No load balancing is possible; one uplink needs to block for all VLANs.
- The CPU is spared; only one instance needs to be computed.

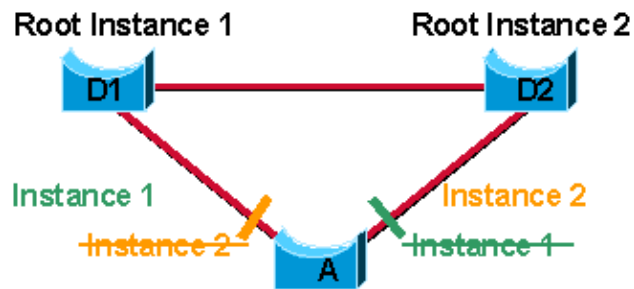
Note: Cisco's implementation enhances the 802.1q in order to support one PVST. This feature behaves exactly as the PVST in the example above. Cisco's per-VLAN BPDUs are tunneled by pure 802.1q bridges.

MST Case

MSTs (IEEE 802.1s) combine the best aspects from both the PVST+ and the 802.1q. The idea is that several VLANs can be mapped to a reduced number of spanning-tree instances because most networks do not need more than a few logical topologies. In the topology described in the first diagram, there are only two different final logical topologies, so only two spanning-tree instances are really necessary. There is no need to run 1,000 instances. If you map half of the 1,000 VLANs to a different spanning-tree instance, as shown in the following diagram, the following is true:

- The desired load balancing scheme can still be achieved, because half of the VLANs follow one separate instance.

- The CPU is spared by only computing two instances.



From a technical stand point, MST is the best solution. From an end-user's perspective, the only drawbacks associated with migrating to MST are mainly attributed to the fact that MST is a new protocol, and the following issues arise:

- The protocol is more complex than the usual spanning-tree and requires additional training of the staff.
- Interaction with legacy bridges is sometimes challenging. For more information refer, to the Interaction Between MST Regions and the Outside World section of this document.

MST Region

As previously mentioned, the main enhancement introduced by MST is that several VLANs can be mapped to a single spanning-tree instance. This raises the problem of determining what VLAN is to be associated with what instance. More precisely, tagging BPDUs so that receiving devices can identify the instances and the VLANs to which they apply.

The issue is irrelevant in the case of the 802.1q standard, where all instances are mapped to a unique and common instance. In the PVST+ implementation, the association is as follows:

- Different VLANs carry the BPDUs for their respective instance (one BPDU per VLAN).

Cisco's MISTP solved this problem by sending a BPDU for each instance and by including in the BPDU a list of VLANs that it was responsible for. If by error, two switches were misconfigured and had a different range of VLANs associated to the same instance, it was difficult for the protocol to recover properly from this situation.

The IEEE 802.1s committee adopted a much easier and simpler approach by introducing MST regions. Think of a region as the equivalent of Border Gateway Protocol (BGP) Autonomous Systems, which is a group of switches placed under a common administration.

MST Configuration and MST Region

Each switch running MST in the network has a single MST configuration that consists of the following three attributes:

1. An alphanumeric configuration name (32 bytes).
2. A configuration revision number (two bytes).

3. A 4096–element table that associates each of the potential 4096 VLANs supported on the chassis to a given instance.

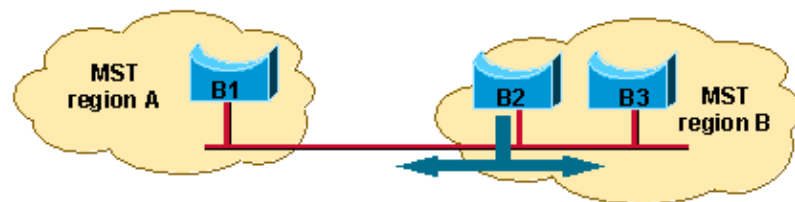
In order to be part of a common MST region, a group of switches must share the same configuration attributes. It is up to the network administrator to properly propagate the configuration throughout the region. Currently, this step is only possible by the means of the Command Line Interface (CLI) or through Simple Network Management Protocol (SNMP). Other methods can be envisioned, as the IEEE specification does not explicitly mention how to accomplish that step.

Note: If for any reason two switches differ on one or more configuration attribute, they are part of different regions. For more information refer to the following Region Boundary section.

Region Boundary

In order to ensure a consistent VLAN–to–instance mapping, it is necessary for the protocol to be able to exactly identify the boundaries of the regions. For that purpose, the characteristics of the region are included in BPDUs. The exact VLANs–to–instance mapping is not propagated in the BPDU, because the switches only need to know whether they are in the same region as a neighbor. Therefore, only a digest of the VLANs–to–instance mapping table is sent, along with the revision number and the name. Once a switch receives a BPDU, it extracts the digest (a numerical value derived from the VLAN–to–instance mapping table through a mathematical function) and compares it with its own computed digest. If the digests differ, the port on which the BPDU was received is at the boundary of a region.

In generic terms, a port is at the boundary of a region if the designated bridge on its segment is in a different region or if it receives legacy 802.1d BPDUs. In the following diagram, the port on B1 is at the boundary of region A, whereas the ports on B2 and B3 are internal to region B.



MST Instances

According to the IEEE 802.1s specification, an MST bridge must be able to handle at least the following two instances:

- One Internal Spanning–Tree (IST).
- One or more Multiple Spanning–Tree Instance(s) (MSTIs).

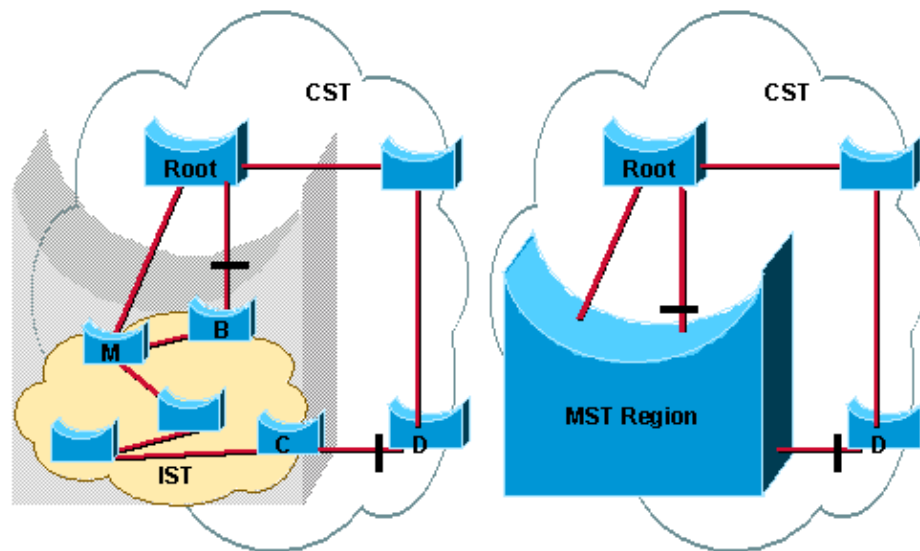
The terminology is evolving, as 802.1s is actually in a pre–standard phase. It is likely these names will change in the final release of 802.1s. Cisco's implementation supports 16 instances: one IST (instance 0) and 15 MSTIs.

IST Instances

In order to clearly understand the role of the IST instance, remember that MST originates from the IEEE. Therefore, MST must be able to interact with 802.1q-based networks, because 802.1q is another IEEE standard. For 802.1q, a bridged network only implements a single spanning-tree (CST). The IST instance is simply an RSTP instance that extends the CST inside the MST region.

The IST instance receives and sends BPDUs to the CST. The IST is capable of representing the entire MST region as a CST virtual bridge to the outside world.

The following are two functionally equivalent diagrams. Notice the location of the different blocked ports. In a typically bridged network, you would expect to see a blocked port between Switches M and B. Instead of blocking on D, you would expect to have the second loop broken by a blocked port somewhere in the middle of the MST region. However, due to the IST, the entire region appears as one virtual bridge that runs a single spanning-tree (CST). This makes it possible to understand that the virtual bridge blocks an alternate port on B. Also, that virtual bridge is on the C to D segment, thus leading Switch D to block its port.

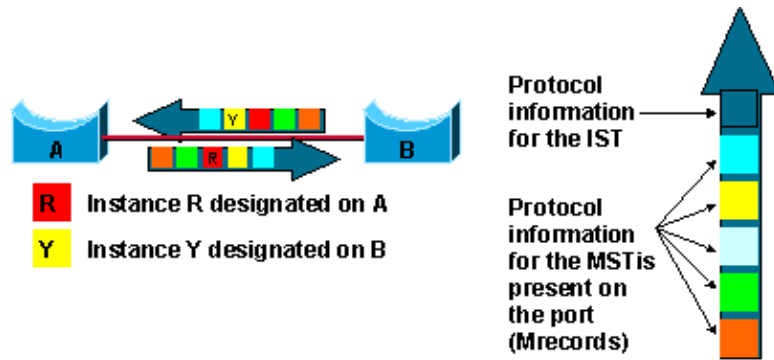


The exact mechanism that makes the region appear as one virtual CST bridge is beyond the scope of this document, and is amply described in the IEEE 802.1s specification. However, by keeping this virtual bridge property of the MST region in mind, the interaction with the outside world is much easier to understand.

MSTIs

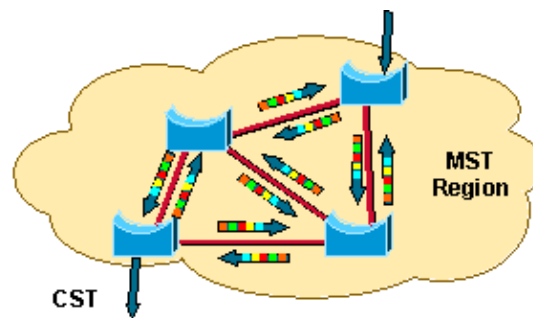
The MSTIs are simple RSTP instances that only exist inside a region. They run the rapid spanning-tree automatically by default, without any extra configuration work. Unlike the IST, MSTIs never interact with the outside of the region. Remember that MST only runs one spanning-tree outside of the region, so except for the IST instance, regular instances inside of the region have no outside counterpart. Additionally, MSTIs do not send BPDUs outside a region, only the IST does.

MSTIs do not send independent individual BPDUs. Inside the MST region, bridges exchange MST BPDUs that can be seen as a normal RSTP BPDU for the IST while containing additional information for each MSTI. The following diagram shows Switches A and B exchanging BPDUs inside an MST region. Each switch only sends one BPDU, but each includes one MRecord per MSTI present on the ports.



Note: In the above diagram, notice that the first information field carried by an MST BDU contains data about the IST. This implies that the IST (instance 0) is always present everywhere inside an MST region. However, the network administrator does not have to map VLANs onto instance 0 and therefore is not a source of concern.

Unlike regular converged spanning-tree topology, both ends of a link may send and receive BPDUs simultaneously. This is because, as shown in the following diagram, each bridge may be designated for one or more instances, and thus needs to transmit BPDUs. As soon as a single MST instance is designated on a port, a BPDUs containing the information for all instances (IST+ MSTIs) is to be sent. The following diagram shows the MST BPDUs being sent inside and outside an MST region:



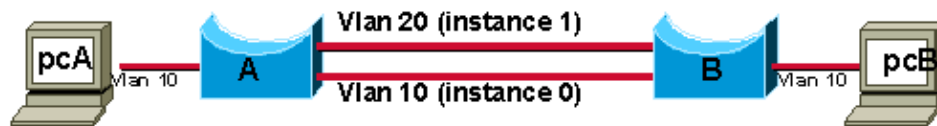
The MRecord contains enough information (mostly root bridge and sender bridge priority parameters) for the corresponding instance to calculate its final topology. It does not need any timer-related parameters such as hello time, forward delay, and max age that are typically found in a regular IEEE 802.1d or 802.1q CST BPDUs. The only instance in the MST region to use these parameters is the IST; the hello time determines how frequently BPDUs are sent and the forward delay parameter is mainly used when rapid transition is not possible (remember that rapid transitions do not occur on shared links). As MSTIs depend on the IST to transmit their information, they do not need those timers.

Common Misconfigurations

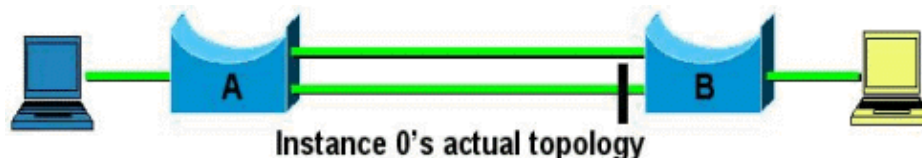
The independence between instance and VLAN is a new concept that implies careful configuration planning. The following section illustrates some common pitfalls and how to avoid them.

IST Instance is Active on All Ports, Whether Trunk or Access

The following diagram shows Switches A and B connected using access ports each located in different VLANs. VLAN 10 and VLAN 20 are mapped to different instances. VLAN 10 is mapped to instance 0, while VLAN 20 is mapped to instance 1.



This configuration results in pcA not being able to send frames to pcB. Issuing the **show** command reveals that Switch B is blocking its link to Switch A in VLAN 10, as shown in the following diagram:



How is that possible in such a simple topology, with no apparent loop?

This issue is explained by the fact that MST information is conveyed using only one BPDU (IST BPDU), regardless of the number of internal instances. Individual instances do not send individual BPDUs. When Switch A and Switch B exchange spanning-tree protocol (STP) information for VLAN 20, they send an IST BPDU with an MRecord for instance 1 because that is where VLAN 20 is mapped. However, because it is an IST BPDU, it also contains information for instance 0. This means that the IST instance is active on all ports inside an MST region, whether these ports carry VLANs mapped to the IST instance or not.

The following diagram shows the logical topology of the IST instance:



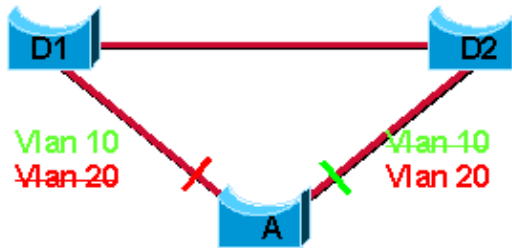
Switch B receives two BPDUs for instance 0 from Switch A (one on each port). It is clear that Switch B has to block one of its ports in order to avoid a loop.

The preferred solution is to avoid mapping VLANs to the IST instance by using one instance for VLAN 10 and another instance for VLAN 20.

An alternative would be to carry those VLANs mapped to the IST on all links (allow VLAN 10 on both ports, as in this previous diagram)

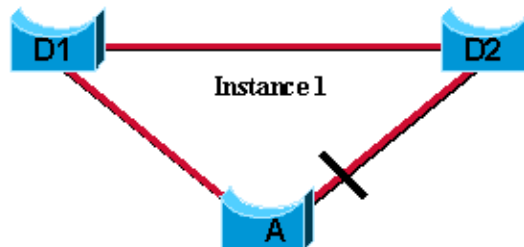
Two VLANs Mapped to the Same Instance Will Block the Same Ports

Remember that VLAN no longer means spanning-tree instance. The topology is determined by the instance, regardless of the VLANs mapped to it. The following diagram shows a problem that is a variant of the one discussed in the above section:



Suppose that VLANs 10 and 20 are both mapped to the same instance (instance 1). The network administrator wants to restrict traffic on the uplink trunks from Switch A to distribution Switches D1 and D2 by manually pruning VLAN 10 on one uplink and VLAN 20 on the other (trying to achieve a topology as described in the above diagram). Shortly after having done that, the network administrator notices that users in VLAN 20 have lost connectivity to the network.

This is a typical misconfiguration problem. VLANs 10 and 20 are both mapped to instance 1, meaning there is only one logical topology for both VLANs. Load-sharing cannot be achieved, as shown in the following diagram:



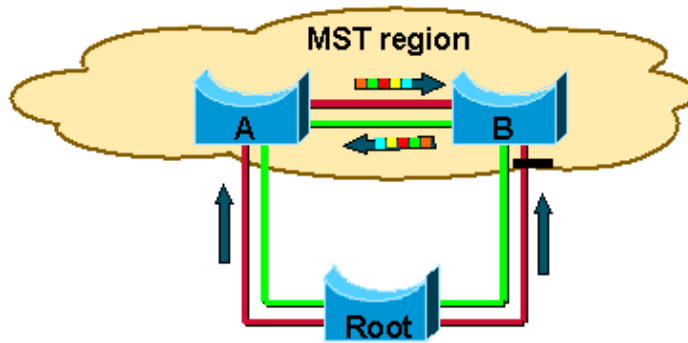
Because of the manual pruning, VLAN 20 is only allowed on the blocked port, which explains the loss of connectivity. To achieve load balancing, the network administrator should have mapped VLAN 10 and 20 to two different instances.

A simple rule to follow to steer clear of this problem is to never manually prune VLANs off a trunk. If you decide to remove some VLANs off a trunk, remove all the VLANs mapped to a given instance together. Never remove an individual VLAN from a trunk without removing all the VLANs that are mapped to the same instance.

Interaction Between the MST Region and the Outside World

When migrating to an MST network, the administrator is likely to have to deal with interoperability issues between MST and legacy protocols. MST seamlessly interoperates with standard 802.1q CST networks, however, only a handful of networks are based on the 802.1q standard because of its single spanning-tree restriction. Cisco released PVST+ at the same time as support for 802.1q was announced. Cisco also provides an efficient yet simple compatibility mechanism between MST and PVST+. This mechanism will be explained in the following paragraphs.

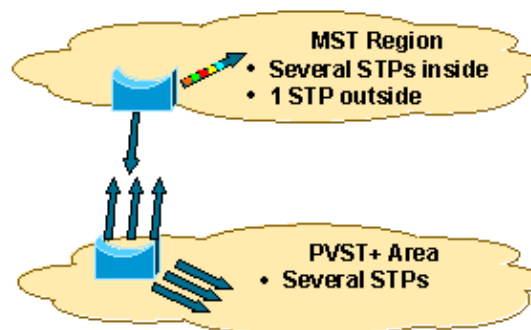
The first property of an MST region is that at the boundary ports no MSTI BPDUs are sent out, only IST BPDUs are. Internal instances (MSTIs) always automatically follow the IST topology at boundary ports, as shown in the following diagram:



In the above diagram, assume VLANs 10–50 are mapped to the green instance, which is an internal instance (MSTI) only. The red links represent the IST, and therefore they also represent the CST. VLANs 10–50 are allowed everywhere in the topology. BPDUs for the green instance are not sent out of the MST region. This does not mean that there is a loop in VLANs 10–50. MSTIs follow the IST at the boundary ports, and the boundary port on Switch B will also block traffic for the green instance.

Switches running MST are able to automatically detect PVST+ neighbors at boundaries. They can do so by detecting that multiple BPDUs are received on different VLANs of a trunk port for the instance .

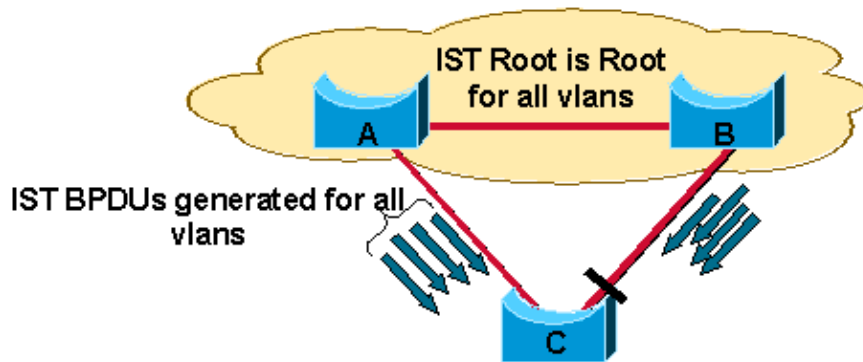
The following diagram shows an interoperability issue. An MST region only interacts with one spanning–tree (the CST) outside of the region. However, PVST+ bridges run one spanning–tree algorithm (STA) per VLAN, and as a result, send one BPDU on each VLAN every two seconds. The boundary MST bridge does not expect to receive that many BPDUs. It either expects to receive one or to send one, depending on whether it is the root of the CST or not.



Cisco developed a mechanism to address the problem shown in the above diagram. A possibility would have consisted in tunneling the extra BPDUs sent by PVST+ bridges across the MST region. However, this solution has proven to be too complex and potentially dangerous when first implemented in the MISTP. A simpler approach was created. The MST region simulates a PVST+ neighbor by replicating the IST BPDU on all the VLANs. This solution implies a few constraints that are discussed in the following section.

Recommended Configuration

As the MST region now replicates the IST BPDUs on every VLAN at the boundary, each PVST+ instance will hear a BPDU from the IST root (this implies the root is located inside the MST region). Cisco recommends that the IST root have a better priority than any other bridge in the network so that the IST root becomes the root for all the different PVST+ instances, as shown in the following diagram:

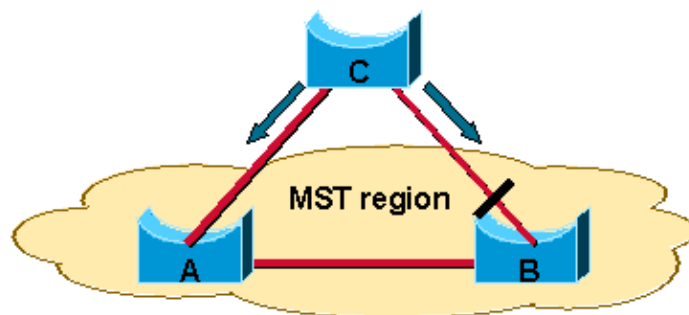


In the above diagram, Switch C is a PVST+ redundantly connected to an MST region. The IST root is the root for all PVST+ instances existing on Switch C. As a result, Switch C blocks one of its uplinks in order to prevent loops. In this particular case, interaction between PVST+ and the MST region is optimal for the following reasons:

- Switch C's uplink ports costs can be tuned to achieve load balancing of the different VLANs across the uplinks' ports (because Switch C runs one spanning-tree per VLAN, it is able to choose which uplink port will block on a per-VLAN basis).
- Uplink fast can be used on Switch C to achieve fast convergence in case of an uplink failure.

Alternate Configuration (Not Recommended)

Another possibility is to have the IST region be the root for absolutely no PVST+ instance. This means that all PVST+ instances have a better root than the IST instance, as shown in the following diagram:



This case corresponds to a PVST+ core and an MST access or distribution layer, a rather infrequent scenario. Establishing the root bridge outside the region has the following drawbacks compared to the previously recommended configuration:

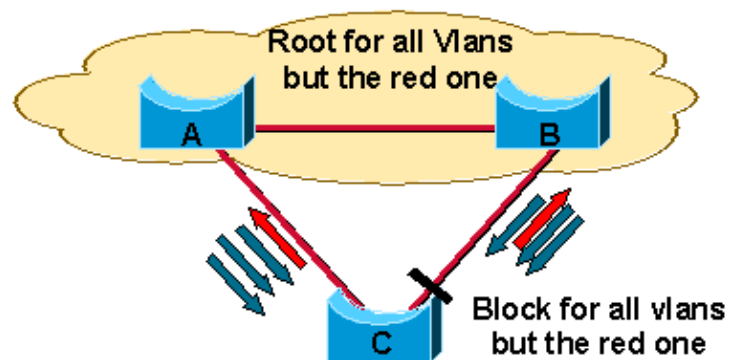
- An MST region only runs one spanning-tree instance that interacts with the outside world. This basically means that a boundary port can only be blocking or forwarding for all VLANs. In other terms, there is no load balancing possible between the region's two uplinks leading to Switch C. The uplink on Switch B, for the instance, will be blocking for all VLANs while Switch A will be forwarding for all VLANs.
- This configuration still allows for fast convergence inside the region. If the uplink on Switch A fails, a fast switch over to an uplink on a different switch needs to be achieved. While the way the IST behaves inside the region in order to have the whole MST region resemble a CST bridge was not discussed in details, you can imagine that a switchover across a region will never be as efficient as a

switchover on a single bridge.

Invalid Configuration

While providing easy and seamless interoperability between MST and PVST+, the PVST+ emulation mechanism implies that any configuration other than the two previously mentioned ones is invalid. The following are the basic rules that must be followed to get a successful MST/PVST+ interaction:

1. If the MST bridge is the root, it must be the root for *all* VLANs.
2. If the PVST+ bridge is the root, it must be the root for *all* VLANs (including the CST, which always runs on VLAN 1, regardless of the native VLAN, when running PVST+).
3. The simulation will fail and produce an error message if the MST bridge is the root for the CST, while the PVST+ bridge is the root for one or more other VLANs. A failed simulation puts the boundary port in root inconsistent mode.



In the above diagram, Bridge A in the MST region is the root for all three PVST+ instances except one (the red VLAN). Bridge C is the root of the red VLAN. Suppose that the loop created on the red VLAN, where Bridge C is the root, becomes blocked by Bridge B. This would mean that Bridge B would be designated for all VLANs except the red one. An MST region is not able to do that. A boundary port can only be blocking or forwarding for all VLANs because the MST region is only running one spanning-tree with the outside world. Thus, when Bridge B detects a better BPDU on its boundary port, it invokes the BPDU guard to block this port. The port is placed in the root inconsistent mode. The exact same mechanism will also lead Bridge A to block its boundary port. Connectivity is lost, however, a loop-free topology is preserved even in the presence of such a misconfiguration.

Note: As soon as a boundary port produces a root inconsistent error, start investigating whether a PVST+ bridge is attempting to become the root for some VLANs.

Migration Strategy

The first step in the migration to 802.1s/w is to properly identify point-to-point and edge ports. Ensure all switch-to-switch links on which a rapid transition is desired, are full-duplex. Edge ports are defined through the portfast feature. Carefully decide how many instances will be needed in the switched network, remembering that an instance translates to a logical topology. Decide what VLANs to map onto those instances, and carefully select a root and a backup root for each instance. Choose a configuration name and a revision number that will be common to all switches in the network. Cisco recommends placing as many switches as possible into a single region; there is no advantage in segmenting a network in separate regions. Avoid mapping any VLANs onto instance 0. Start by migrating the core first by changing the STP type to

MST, and work your way down to the access switches. MST can interact with legacy bridges running PVST+ on a per-port basis, so it is not a problem to mix both types of bridges if interactions are clearly understood. Always try to keep the root of the CST/IST inside the region. If you are interacting with a PVST+ bridge through a trunk, ensure the MST bridge is the root for all VLANs allowed on that trunk.

Conclusion

Switched networks must fulfill stringent robustness, resiliency, and high-availability requirements. With growing technologies such as voice and video over IP, fast convergence around link or components failures is no longer a desirable characteristic, it is a must. However, until recently, redundant switched networks had to rely on the relatively sluggish 802.1d STP to achieve those goals. This often turned out to be the network administrator's most challenging task, as tuning the protocol timers was the only way to get a few seconds of the protocol, but often at the detriment of the network's health. Cisco has released many 802.1d STP augmentations such as uplink fast, backbone fast and portfast, features that paved the way toward faster spanning-tree convergence. Cisco also answered large Layer 2-based networks' scalability issues by developing the MISTP. The IEEE recently decided to incorporate most of these concepts into two standards: 802.1w (RSTP) and 802.1s (MST). Using these new protocols, convergence times in the low hundreds of milliseconds can be expected while scaling to thousands of VLANs. Cisco remains the leader in the industry by offering the two protocols along with proprietary augmentations in order to facilitate the migration and interoperability with legacy bridges.

Tools Information

For additional resources, refer to Cisco TAC Tools for LAN Technologies.

Related Information

Related Topics

- [Understanding Rapid Spanning-Tree Protocol \(802.1w\)](#)

Additional Documentation

- [LAN Technologies Technical Tips](#)
 - [LAN Technologies Top Issues](#)
-

All contents are Copyright © 1992--2002 Cisco Systems, Inc. All rights reserved. Important Notices and Privacy Statement.

Updated: Jun 14, 2002

Document ID: 24248
