Understand the Multiple Spanning Tree Protocol (802.1s)

Contents

Introduction

Prerequisites

- Requirements
- Components Used

Background Information

Where to Use MST

- PVST+ Case
- Standard 802.1q Case
- MST Case

MST Region

MST Configuration and MST Region

Region Boundary

- MST Instances
- IST Instances
- MSTIs
- Common Misconfigurations

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

Two VLANs Mapped to the Same Instance Block the Same Ports

Interaction Between the MST Region and the Outside World

Recommended Configuration

- Alternate Configuration (Not Recommended)
- Invalid Configuration

Migration Strategy

Conclusion

Related Information

Introduction

This document describes the features and configurations for the Multiple Spanning Tree Protocol (802.1s).

Prerequisites

Requirements

Cisco recommends that you have knowledge of these topics:

- A familiarity with Rapid STP (RSTP) (802.1w)

Components Used

This document is not restricted to specific software and hardware versions.

The information in this document was created from the devices in a specific lab environment. All of the devices used in this document started with a cleared (default) configuration. If your network is live, ensure that you understand the potential impact of any command.

Background Information

Multiple Spanning Tree (MST) is an IEEE standard inspired from the Cisco proprietary Multiple Instances Spanning Tree Protocol (MISTP) implementation. This table shows the support for MST in various Catalyst switches:

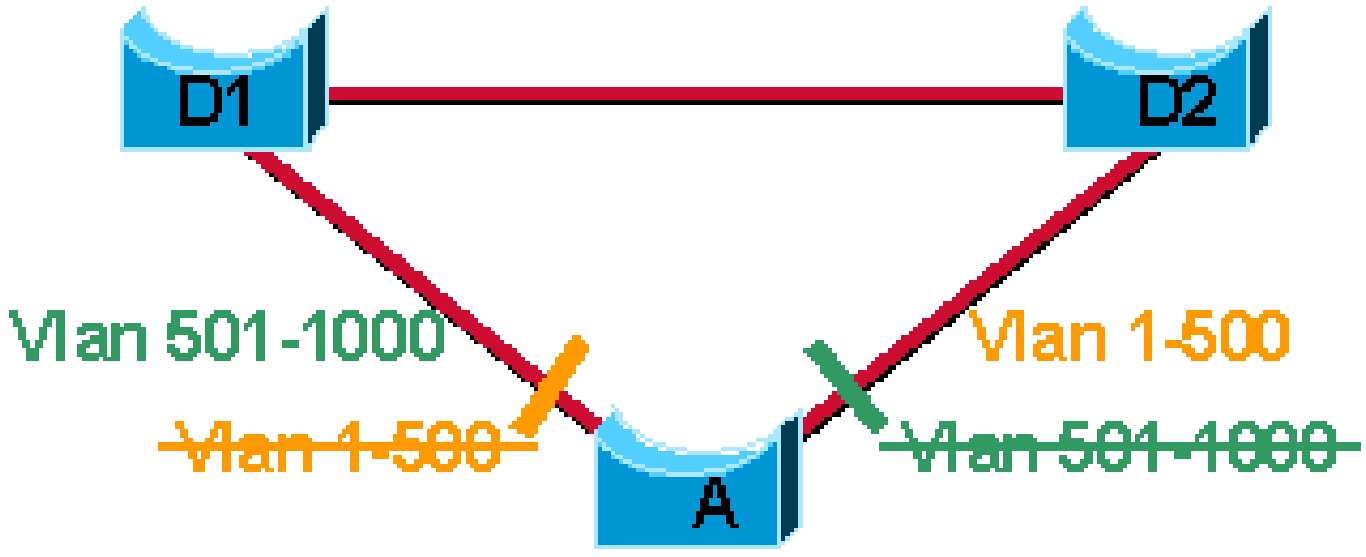
Catalyst Platform	MST with RSTP
Catalyst 2900 XL and 3500 XL	Not Available
Catalyst 2950 and 3550	Cisco IOS® 12.1(9)EA1
Catalyst 2955	All Cisco IOS versions
Catalyst 2948G-L3 and 4908G-L3	Not Available
Catalyst 4000 and 4500 (Cisco IOS)	12.1(12c)EW
Catalyst 5000 and 5500	Not Available
Catalyst 6000 and 6500 (Cisco IOS)	12.1(11b)EX, 12.1(13)E, 12.2(14)SX
Catalyst 8500	Not Available

For more information on RSTP (802.1w), refer to Understanding Rapid Spanning Tree Protocol (802.1w).

Where to Use MST

This diagram shows a common design that features 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.



Access Switch A with 1000 VLANs Redundantly Connected to Switches D1 and D2

These sections are example cases where different types of STP are used on this setup:

PVST+ Case

In a Cisco Per-VLAN Spanning Tree (PVST+) environment, the spanning tree parameters are tuned so that half of the VLANs forward on each Uplink trunk.

In order to easily achieved this, elect Bridge D1 to be the root for VLANs 501 through 1000, and Bridge D2 to be the root for VLANs 1 through 500. These statements are true for this configuration:

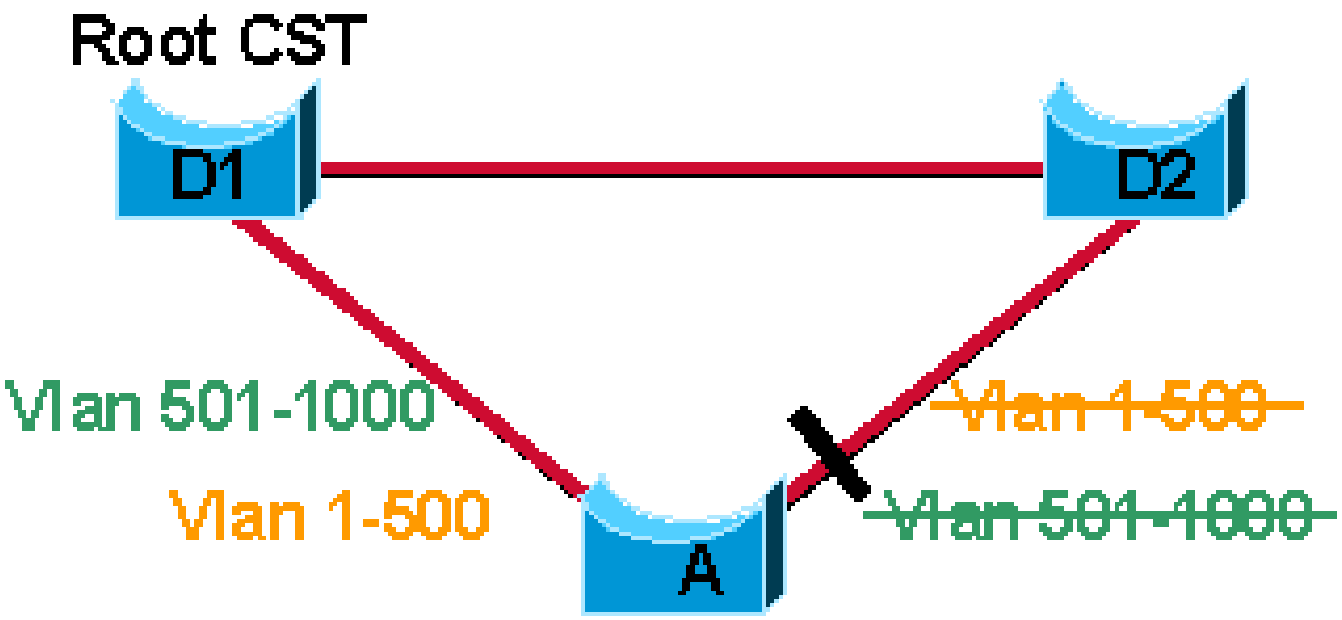
- In this case, optimum load balancing results.
- One spanning tree instance for each VLAN is maintained, which means 1000 instances for only two different final logical topologies.

This considerably wastes CPU cycles for all of the switches in the network (in addition to the bandwidth used for each instance to send its own Bridge Protocol Data Units (BPDUs)).

Standard 802.1q Case

The original IEEE 802.1q standard defines much more than simply trunking. This standard 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 ofthis next diagram, the result resembles the diagram shown here:



Common Spanning Tree (CST) Applied to Network

In a network that runs the CST, these statements 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: The Cisco implementation enhances the 802.1q in order to support one PVST. This feature behaves exactly as the PVST in this example. The Cisco per-VLAN BPDUs are tunneled by pure 802.1q bridges.

MST Case

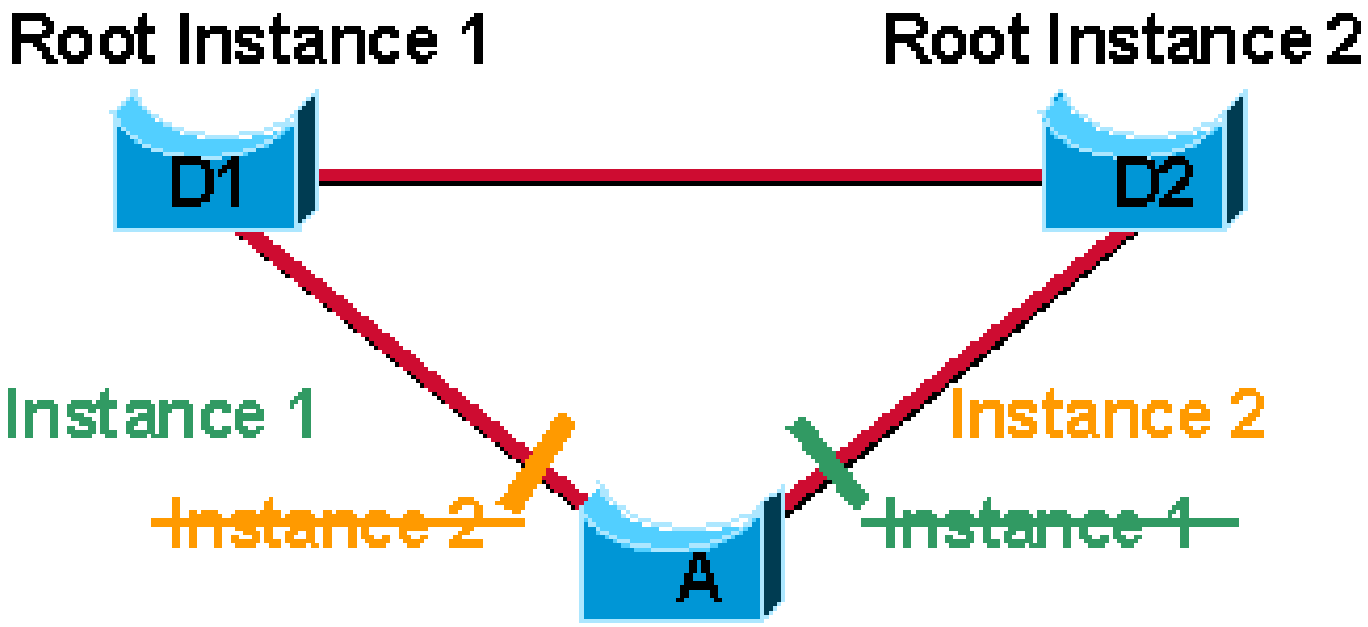
MSTs (IEEE 802.1s) combine the best aspects from both the PVST+ and the 802.1q.

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 1000 instances. If you map half of the 1000 VLANs to a different spanning tree instance, as shown in this diagram, these statements are true:

- The desired load balancing scheme can still be achieved because half of the VLANs adhere to one separate instance.
- The CPU is spared because only two instances are computed.



Map Half of the 1000 VLANs to a Different Spanning Tree Instance

From a technical standpoint, MST is the best solution. From an end-user perspective, the main drawbacks associated with a migration to MST are:

- The protocol is more complex than the usual spanning tree and requires additional training of the staff.
- Interaction with legacy bridges can be a challenge. 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 how to determine which VLAN is to be associated with which instance; more precisely, how to tag BPDUs so that the receiving devices can identify the instances and the VLANs to which each device applies.

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

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

The Cisco MISTP sent a BPDU for each instance, with a list of VLANs that the BPDU was responsible for, in order to solve this problem.

If by error, two switches were not configured correctly 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 that introduced 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 that runs MST in the network has a single MST configuration that consists of these 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, the switches are part of different regions. For more information refer to the next section, **Region Boundary**.

Region Boundary

In order to ensure 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 the BPDUs. The exact VLANs-to-instance

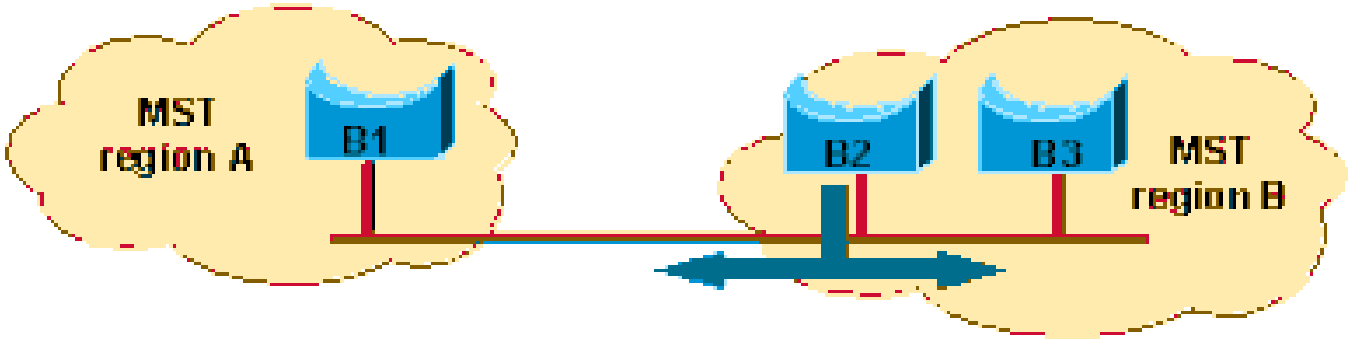
mapping is not propagated in the BPDU, because the switches only need to know whether they are in the same region as a neighbor.

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, the switch extracts the digest (a numerical value derived from the VLAN-to-instance mapping table through a mathematical function) and compares this digest 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 this 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

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

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

The terminology continues to evolve, as 802.1s is actually in a pre-standard phase. It is likely these names can change in the final release of 802.1s. The Cisco 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 can represent the entire MST region as a CST virtual bridge to the outside world.

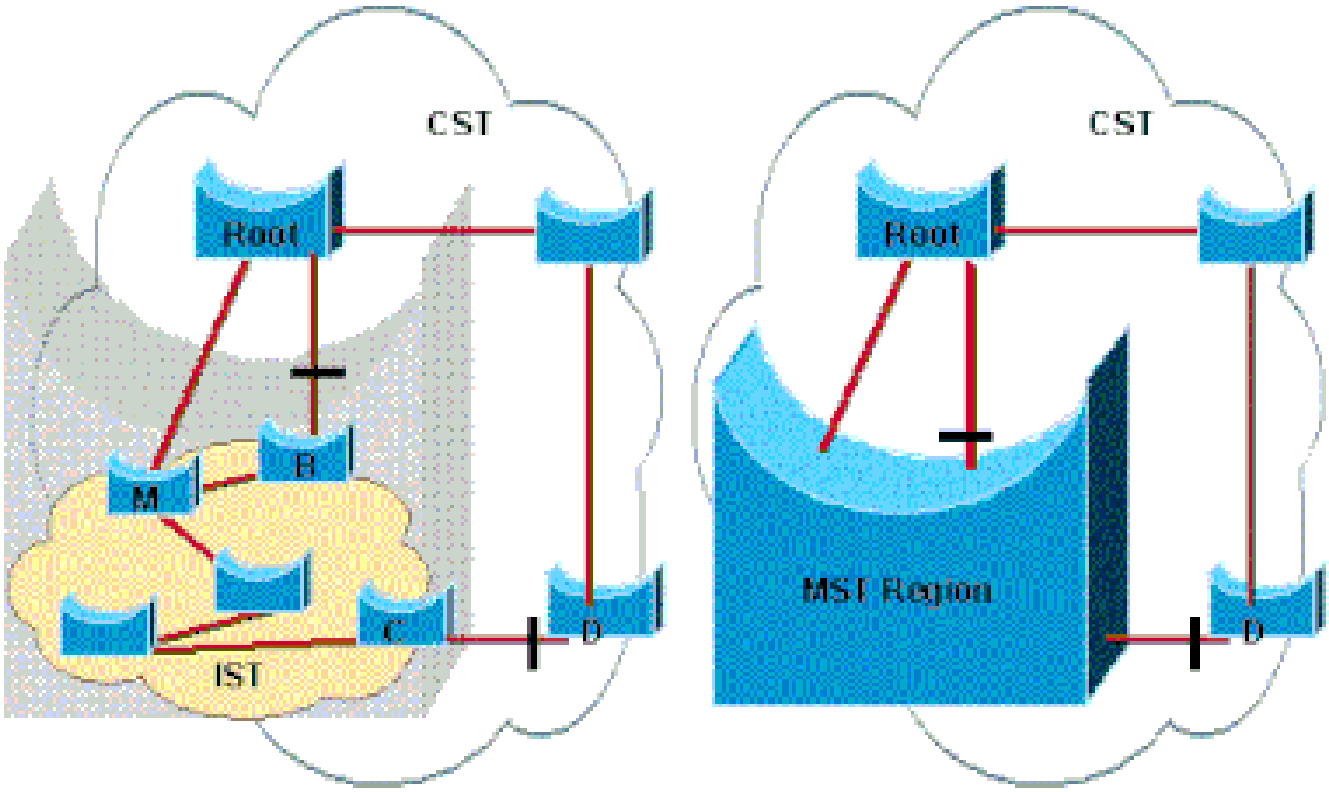
These are two functionally equivalent diagrams. Notice the location of the different blocked ports. In a typically bridged network, you expect to see a blocked port between Switches M and B.

Instead of blocking on D, you 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 and leads 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 but is amply described in the IEEE 802.1s specification.

However, if you keep this virtual bridge property of the MST region in mind, the interaction with the outside world is much easier to understand.

MSTIs

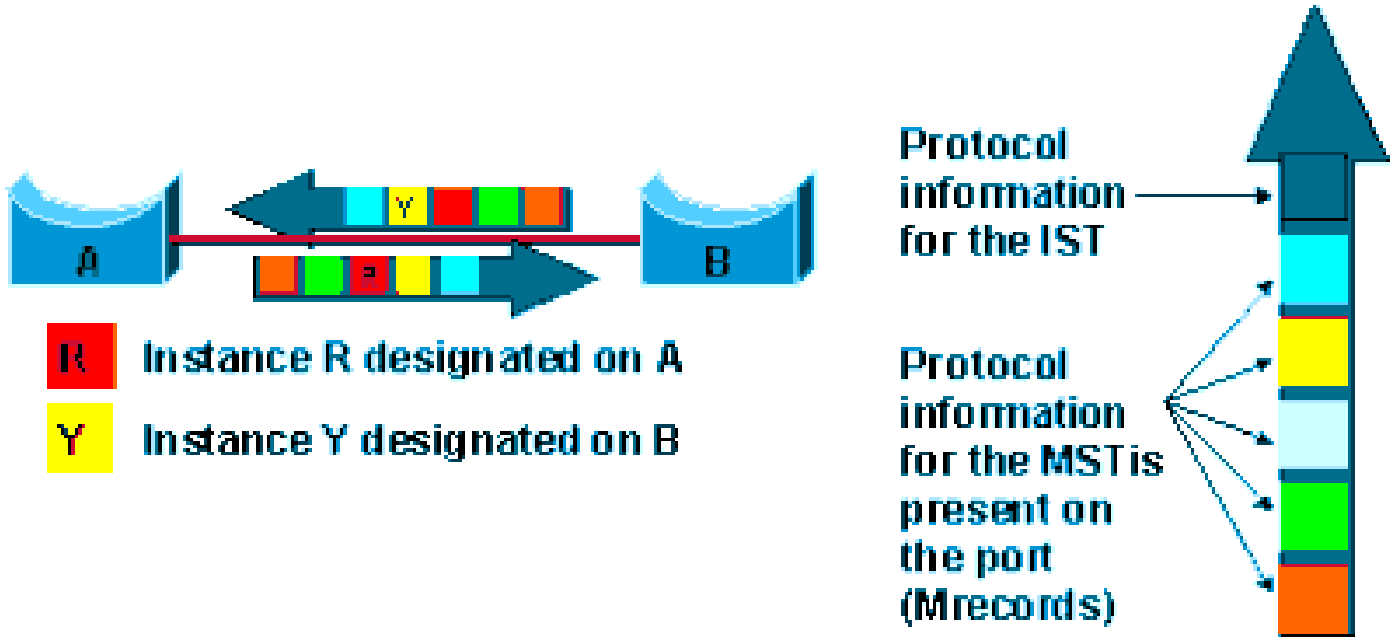
The MSTIs are simple RSTP instances that only exist inside a region. These instances run the RSTP automatically by default, without any extra configuration work.

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 normal RSTP BPDUs for the IST and also contain additional information for each MSTI.

This diagram shows a BPDU exchange between Switches A and B inside an MST region. Each switch only sends one BPDU, but each includes one MRecord per MSTI present on the ports.



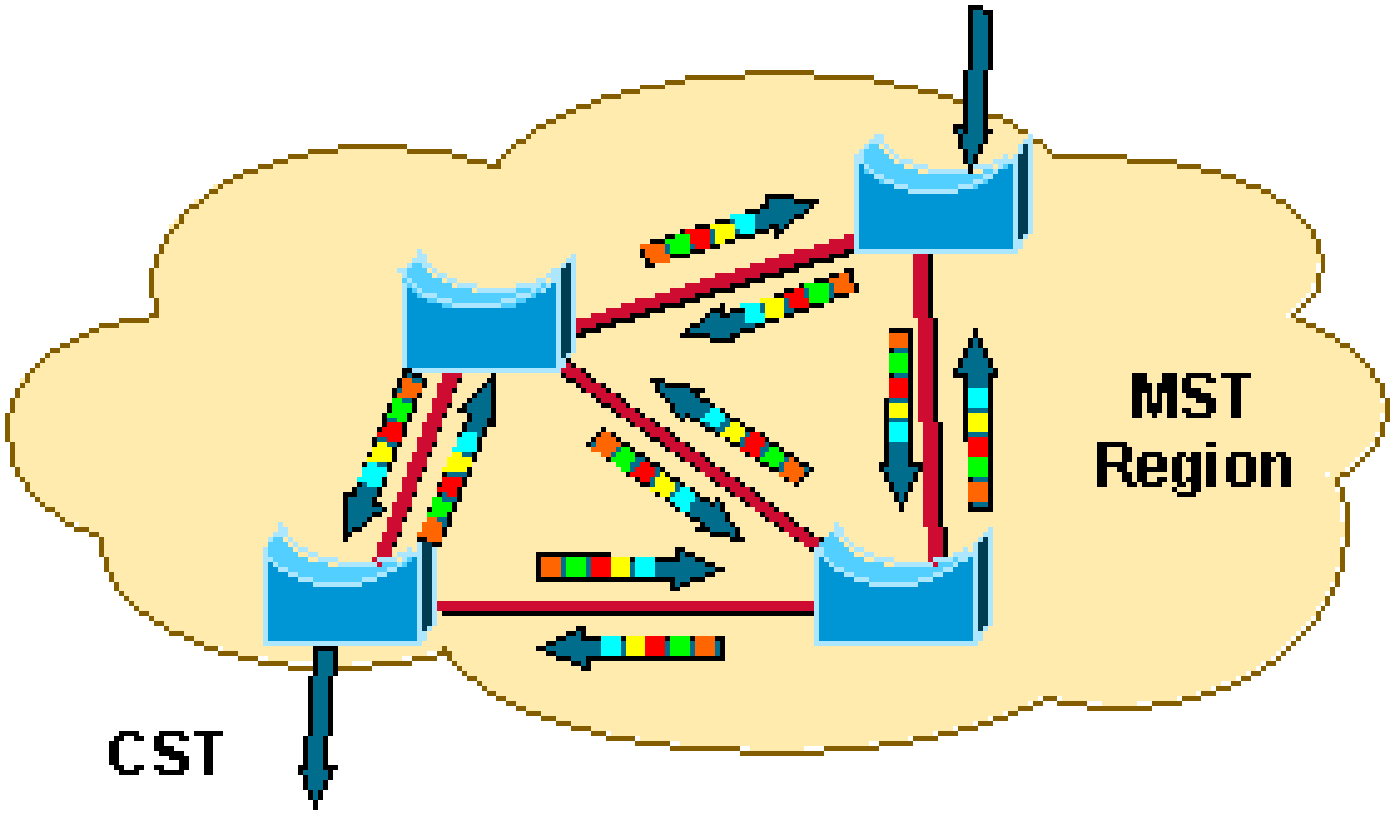
Note: In this diagram, notice that the first information field carried by an MST BPDU contains data about the IST. This implies that the IST (instance 0) is always present everywhere inside an MST region. However, the network administrator does not have to map VLANs onto instance 0, and therefore this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

Unlike regular converged spanning tree topology, both ends of a link can send and receive BPDUs simultaneously.

This is because, as shown in this diagram, each bridge can be designated for one or more instances and needs to transmit BPDUs.

As soon as a single MST instance is designated on a port, a BPDU that contains the information for all instances (IST+ MSTIs) is to be sent.

The diagram shown here demonstrates MST BDPUs sent inside and outside of an MST region:



MST BDPUs Sent Inside and Outside of 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.

The MRecord 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 BPDU.

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).

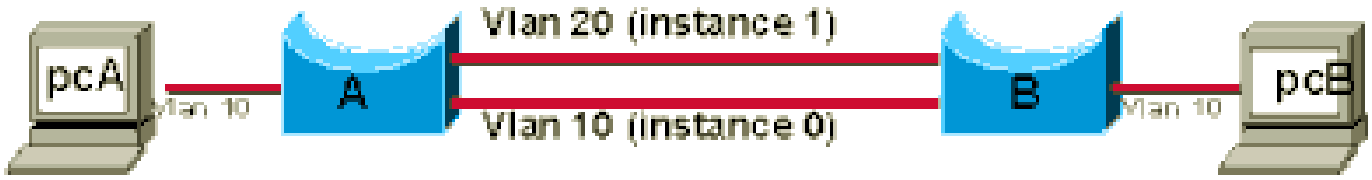
As MSTIs depend on the IST to transmit their information, MSTIs do not need those timers.

Common Misconfigurations

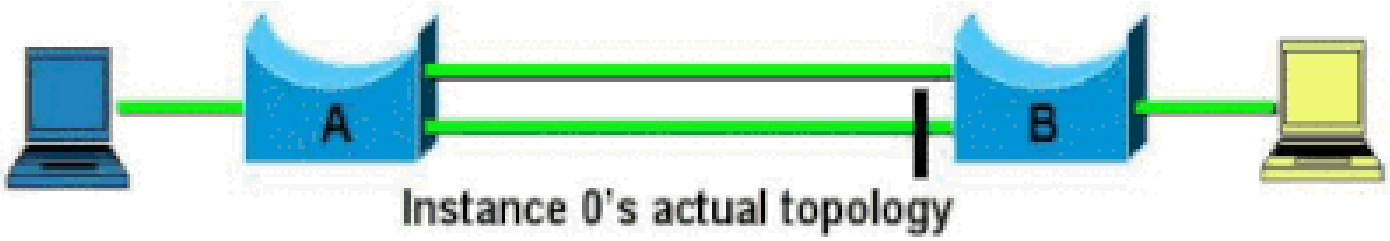
The independence between instance and VLAN is a new concept that implies you must carefully plan your configuration. The IST Instance is Active on All Ports, Whether Trunk or Access section illustrates some common pitfalls and how to avoid them.

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

This diagram shows Switches A and B connected with access ports located in separate 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 's inability to send frames to pcB. The **show** command reveals that Switch B is blocking the link to Switch A in VLAN 10, as shown in this diagram:



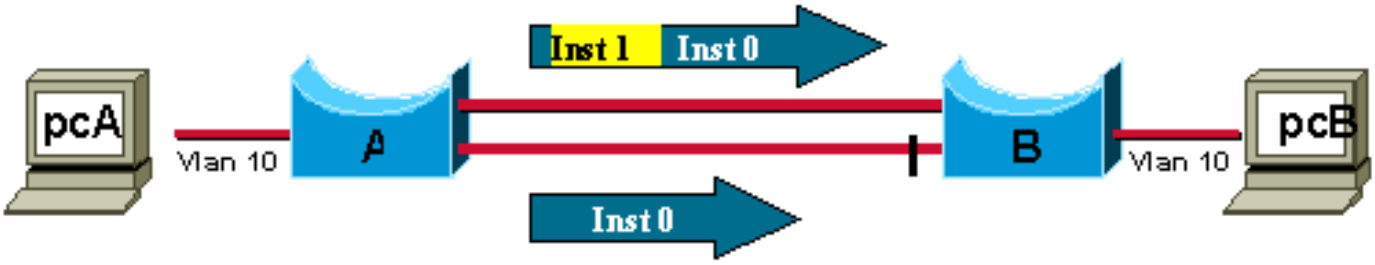
How is that possible in such a simple topology, with no apparent loop?

This issue is explained by the fact that MST information is conveyed with 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 STP information for VLAN 20, the switches 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, this BPDU 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.

This 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.

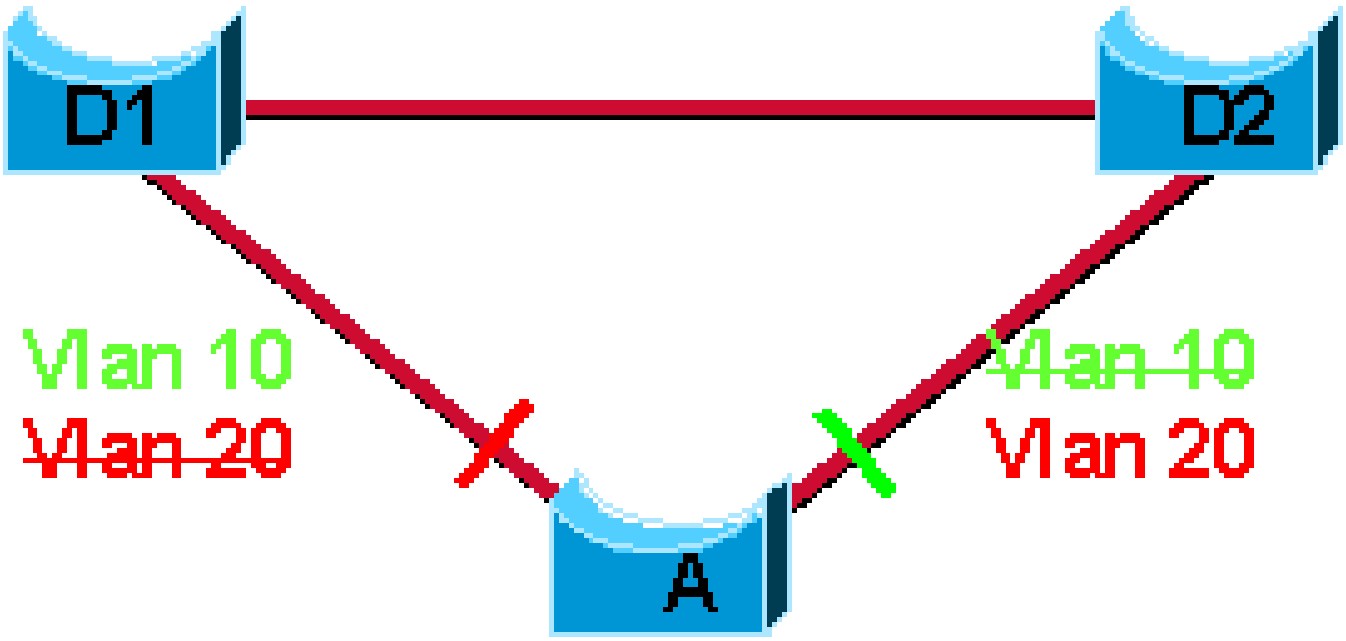
The preferred solution is to use one instance for VLAN 10 and another instance for VLAN 20 to avoid mapping VLANs to the IST instance.

An alternative is to carry those VLANs mapped to the IST on all links (allow VLAN 10 on both ports, as in the next diagram).

Two VLANs Mapped to the Same Instance 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.

This diagram shows a problem that is a variant of the one discussed in the [IST Instance is Active on All Ports, Whether Trunk or Access] section:



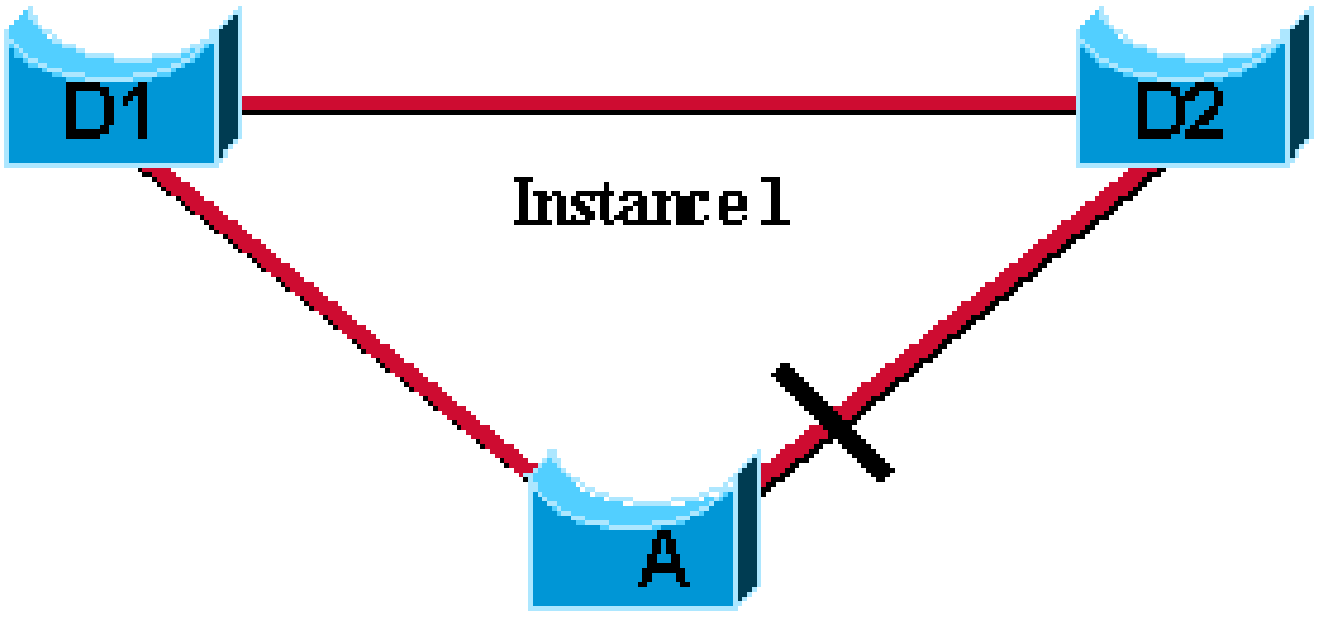
The Topology is Determined by the Instance, Regardless of the VLANs Mapped to It

Suppose that VLANs 10 and 20 are both mapped to the same instance (instance 1).

The network administrator wants to manually prune VLAN 10 on one Uplink and VLAN 20 on the other in order to restrict traffic on the Uplink trunks from Switch A to distribution Switches D1 and D2 (an attempt to achieve a topology as described in the previous diagram).

Shortly after this is completed, 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, which means there is only one logical topology for both VLANs. Load-sharing cannot be achieved, as shown here:



Typical Misconfiguration Problem

Because of the manual pruning, VLAN 20 is only allowed on the blocked port, which explains the loss of connectivity. In order to achieve load balancing, the network administrator must map VLAN 10 and 20 to two different instances.

A simple rule to use 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 and not remove all the VLANs that are mapped to the same instance.

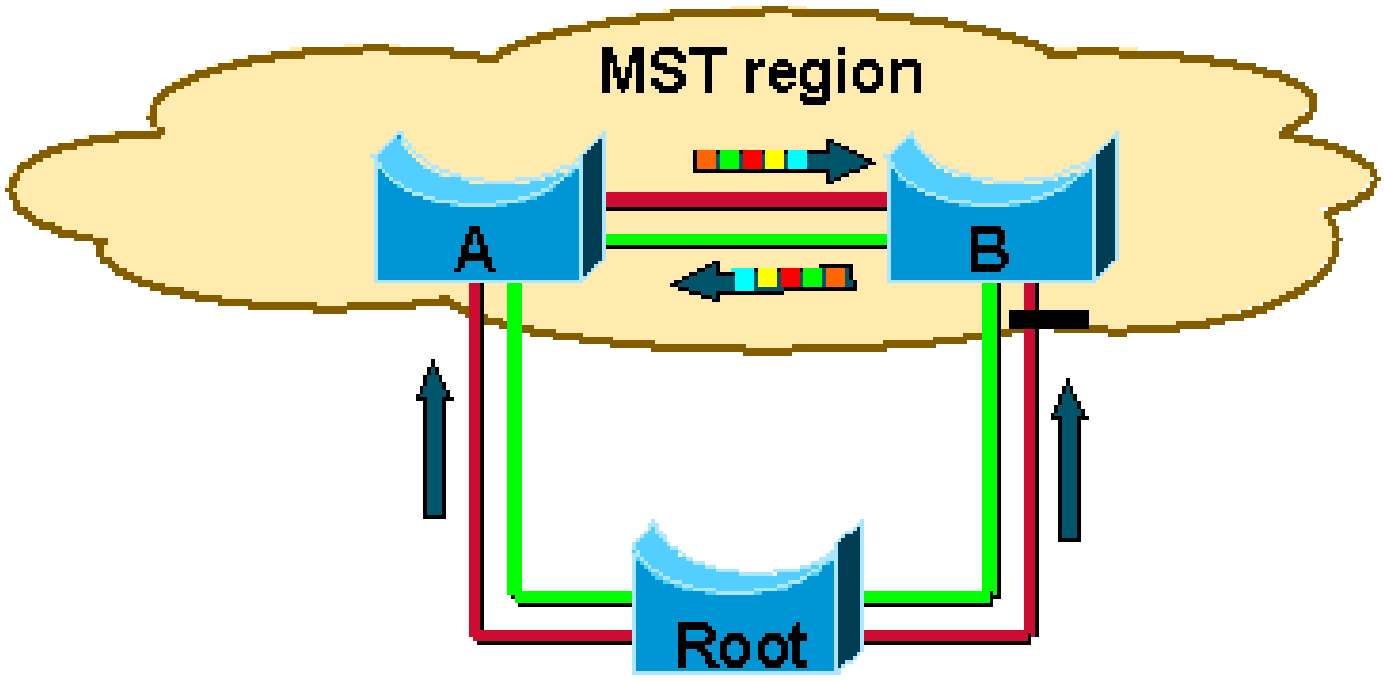
Interaction Between the MST Region and the Outside World

With a migration to an MST network, the administrator is likely to have to deal with interoperability issues between MST and legacy protocols.

MST seamlessly operates 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 is explained later in this document.

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 match the IST topology at boundary ports, as shown in this diagram:



Internal Instances (MSTIs) Always Automatically Match the IST Topology at Boundary Ports

In this diagram, assume VLANs 10 through 50 are mapped to the green instance, which is an internal instance (MSTI) only.

The red links represent the IST, and therefore also represent the CST. VLANs 10 through 50 are allowed everywhere in the topology.

BPDUs for the green instance are not sent out of the MST region.

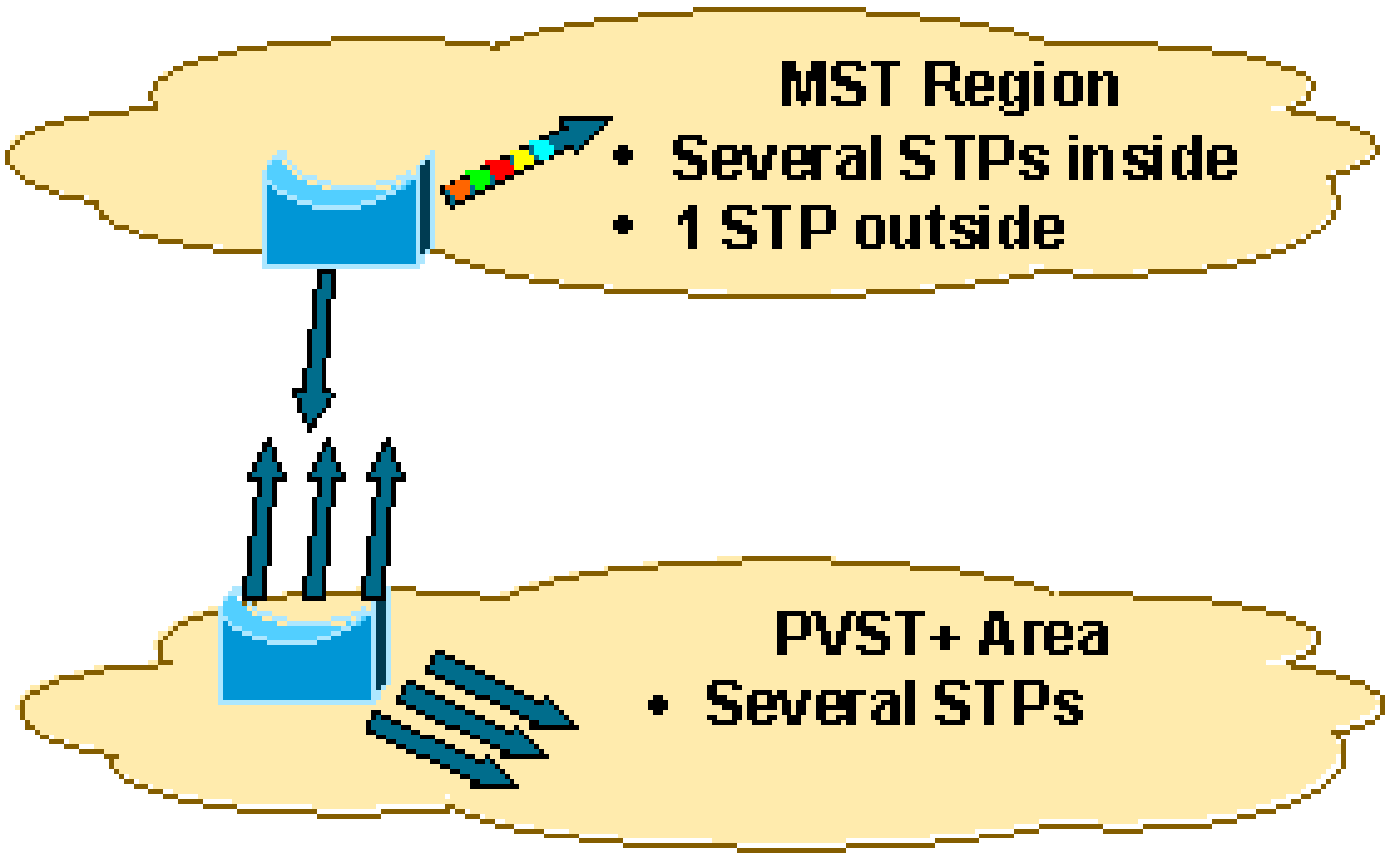
This does not mean that there is a loop in VLANs 10 through 50. MSTIs track the IST at the boundary ports, and the boundary port on Switch B also blocks traffic for the green instance.

Switches that run MST are able to automatically detect PVST+ neighbors at boundaries. These switches are able to detect that multiple BPDUs are received on different VLANs of a trunk port for the instance.

This 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. The MST bridge either expects to receive one or to send one, this depends on whether the bridge is the root of the CST or not.



The MST Bridge Either Expects to Receive One or to Send One

Cisco developed a mechanism to address the problem shown in this diagram. A possibility could have consisted of tunneling the extra BPDUs sent by the PVST+ bridges across the MST region.

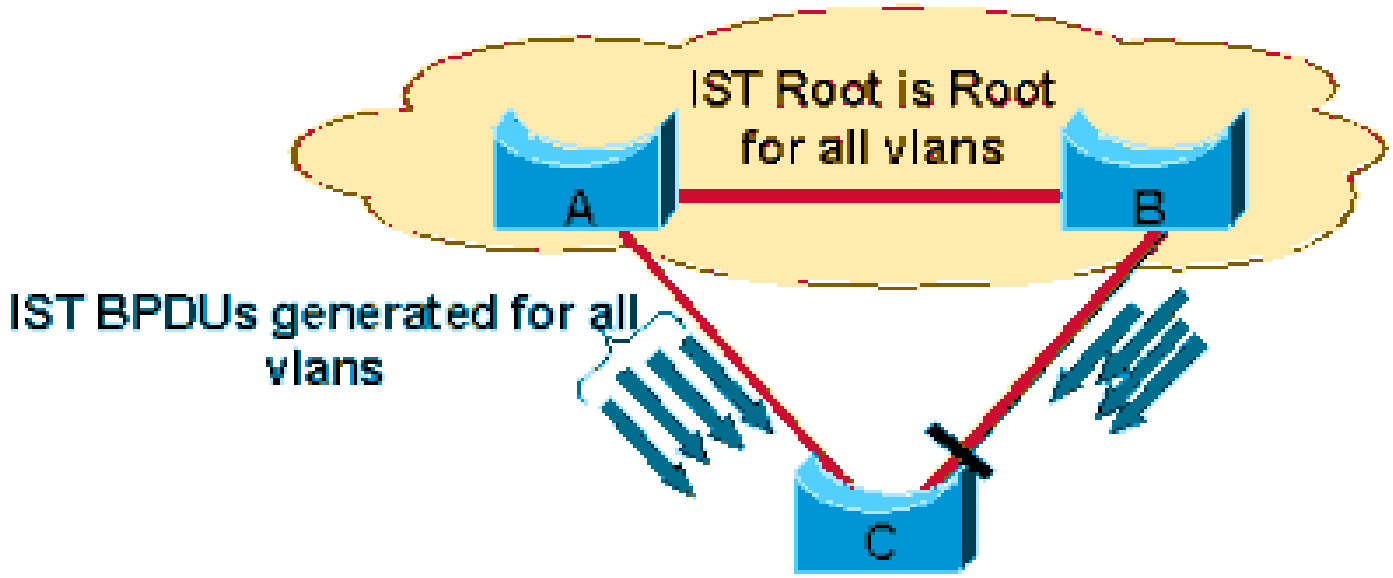
However, this solution provided to be too complex and potentially dangerous when first implemented in the MISTP. A simpler approach was created.

The MST region replicates the IST BPDU on all the VLANs to simulate a PVST+ neighbor. This solution implies a few constraints that are discussed in this document.

Recommended Configuration

As the MST region now replicates the IST BPDUs on every VLAN at the boundary, each PVST+ instance hears a BPDU from the IST root (this implies the root is located inside the MST region).

It is recommended that the IST root have a higher priority than any other bridge in the network so that the IST root becomes the root for all of the different PVST+ instances, as shown in this diagram:



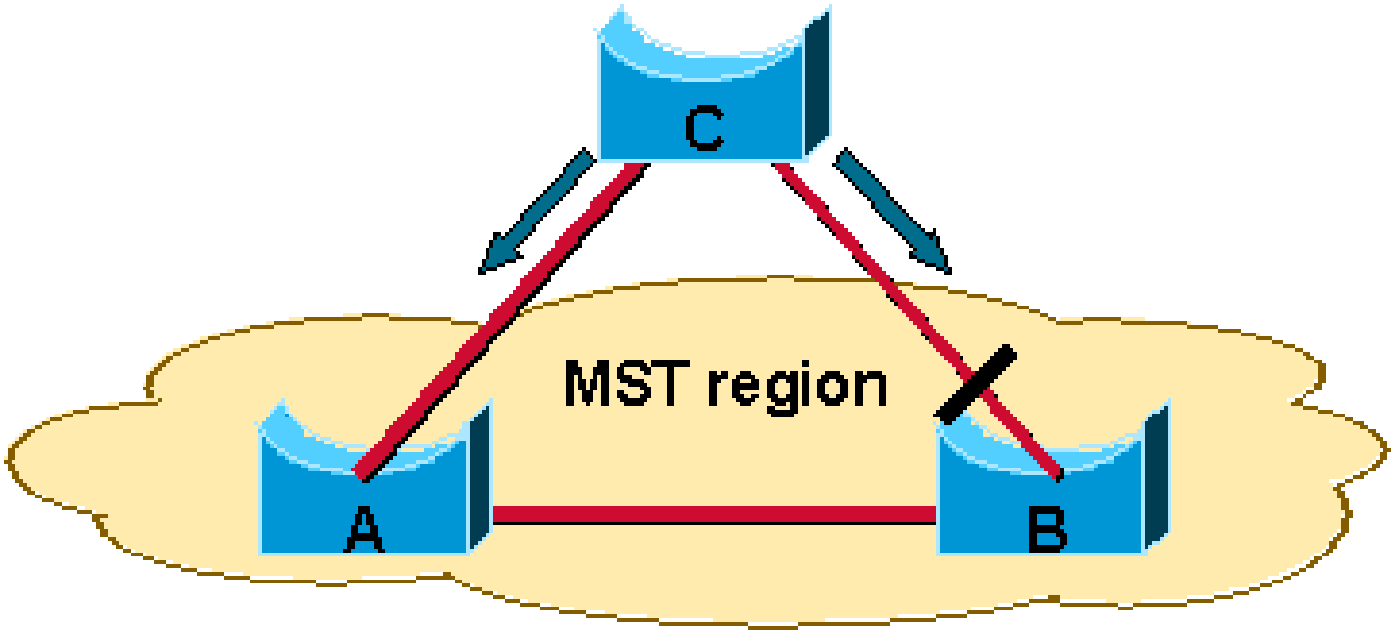
In this diagram, Switch C is a PVST+ redundantly connected to an MST region. The IST root is the root for all PVST+ instances that exist on Switch C.

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 because:

- Switch C 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, this switch is able to choose which Uplink port blocks on a per-VLAN basis).
- UplinkFast 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 this diagram:



All PVST+ Instances Have a Better Root Than the IST Instance

This case corresponds to a PVST+ core and an MST access or distribution layer, a rather infrequent scenario. If you establish the root bridge outside the region, there are these drawbacks as 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 two Uplinks of the region that lead to Switch C. The Uplink on Switch B for the instance can be blocking for all VLANs while Switch A can be forwarding for all VLANs.

- This configuration still allows for fast convergence inside the region. If the Uplink on Switch A fails, a fast switchover 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 detail, you can imagine that a switchover across a region is never as efficient as a switchover on a single bridge.

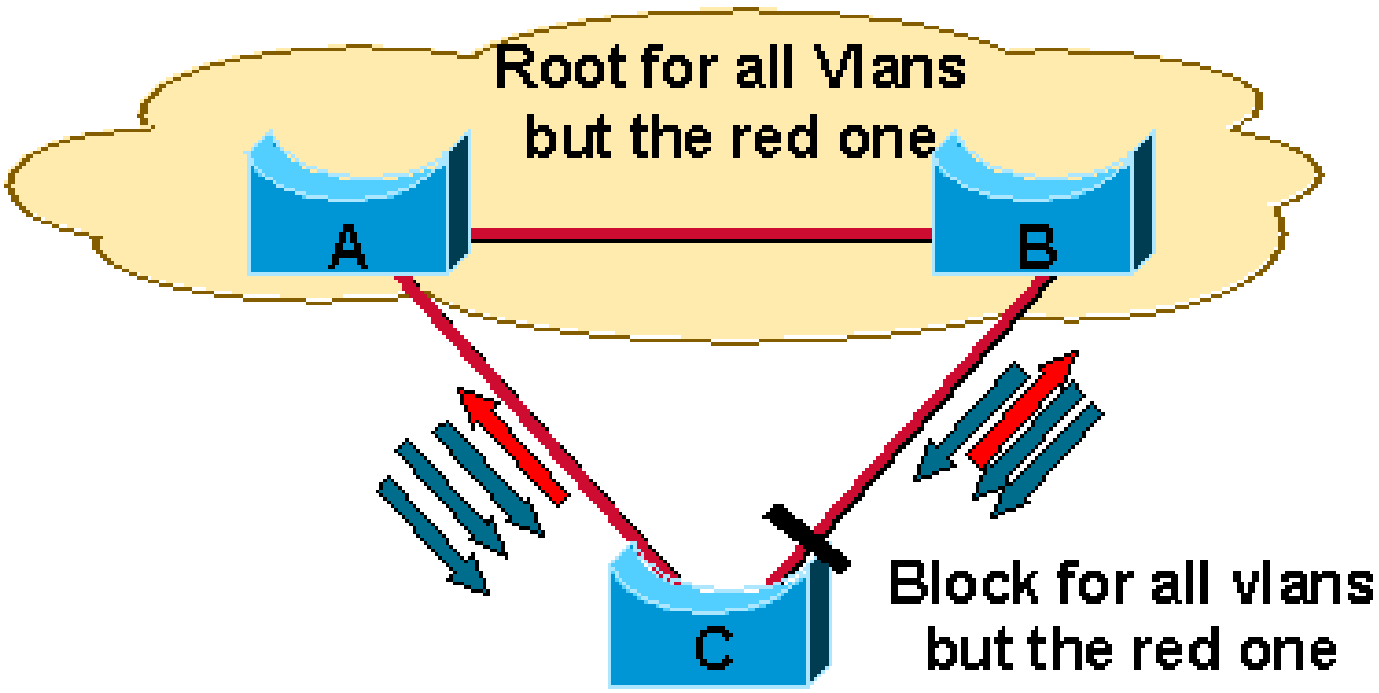
Invalid Configuration

While the PVST+ emulation mechanism provides easy and seamless interoperability between MST and PVST+, this mechanism implies that any configuration other than the two previously mentioned is invalid. These are the basic rules that must be adhered to for a successful MST and PVST+ interaction:

1. If the MST bridge is the root, this bridge must be the root for all VLANs.

2. If the PVST+ bridge is the root, this bridge must be the root for all VLANs (this includes the CST, which always runs on VLAN 1, regardless of the native VLAN, when the CST runs PVST+).

3. The simulation fails and produces 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.



A Failed Simulation Puts the Boundary Port in Root Inconsistent Mode

In this 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 means that Bridge B is 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 only runs one spanning tree with the outside world.

Thus, when Bridge B detects a better BPDU on its boundary port, the bridge invokes the BPDU guard to block this port. The port is placed in the root inconsistent mode.

The exact same mechanism also leads 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, investigate whether a PVST+ bridge has attempted 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 are needed in the switched network, and keep in mind that an instance translates to a logical topology.

Decide what VLANs to map onto those instances, and carefully select a root and a back-up root for each instance.

Choose a configuration name and a revision number that can be common to all switches in the network.

Cisco recommends that you place as many switches as possible into a single region; it is not advantageous to segment a network into separate regions.

Avoid mapping any VLANs onto instance 0. Migrate the core first. Change the STP type to MST and work your way down to the access switches.

MST can interact with legacy bridges that run 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 and IST inside the region. If you interact with a PVST+ bridge through a trunk, ensure the MST bridge is the root for all VLANs allowed on that trunk.

For sample configurations, refer to:

- Configuration Example to Migrate the Spanning Tree from PVST+ to MST

Conclusion

Switched networks must fulfill stringent robustness, resiliency, and high-availability requirements.

With new technologies such as Voice over IP (VoIP) and Video over IP, fast convergence around link or component failures is no longer a desirable characteristic: fast convergence 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 task that challenged the network administrator.

The only way to get a few seconds off the protocol was to tune the protocol timers, but often at the detriment to the health of the network.

Cisco has released many 802.1d STP augmentations such as UplinkFast, BackboneFast and PortFast. These features paved the way toward faster spanning tree convergence.

Cisco also answered large Layer 2 (L2)-based networks' scalability issues with the development of the MISTP. The IEEE recently decided to incorporate most of these concepts into two standards: 802.1w (RSTP) and 802.1s (MST).

With the implementation of these new protocols, you can expect convergence times in the low hundreds of milliseconds and at the same time scale to thousands of VLANs.

Cisco remains the leader in the industry and offers these two protocols along with proprietary augmentations in order to facilitate the migration of and interoperability with legacy bridges.

Related Information

- **Understanding Rapid Spanning Tree Protocol (802.1w)**
- **LAN Switching Technology Support**
- **Cisco Technical Support & Downloads**