Best Practices for Catalyst 4500/4000, 5500/5000, and 6500/6000 Series Switches Running CatOS Configuration and Management

Contents

Introduction

This document discusses the implementation of Cisco Catalyst series switches in your network, specifically the Catalyst 4500/4000, 5500/5000, and 6500/6000 platforms. Configurations and commands are discussed under the assumption that you are running Catalyst OS (CatOS) General Deployment software 6.4(3) or later. Although some design considerations are presented, this document does not cover overall campus design.

Prerequisites

Requirements

This document assumes familiarity with the Catalyst 6500 Series Command Reference, 7.6.

Although references to public online material for further reading are provided throughout the document, these are other foundational and educational references:

- Cisco ISP Essentials— Essential IOS Features Every ISP Should Consider.
- Cisco Network Monitoring and Event Correlation Guidelines
- Gigabit Campus Network Design—Principles and Architecture
- Cisco SAFE: A Security Blueprint for Enterprise Networks

Components Used

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

Conventions

Refer to Cisco Technical Tips Conventions for more information on document conventions.

Background Information

These solutions represent years of field experience from Cisco engineers working with many of our largest customers and complex networks. Consequently, this document emphasizes real-world configurations that make networks successful. This paper offers these solutions:

- Solutions that have statistically the broadest field exposure, and thus the lowest risk.
- Solutions that are simple, trading some flexibility for deterministic results.
- Solutions that are easy to manage and configured by network operations teams.
- Solutions that promote high availability and high stability.

This document is divided into these four sections:

- Basic Configuration— features used by a majority of networks such as Spanning Tree

Protocol (STP) and trunking.

- Management Configuration— design considerations along with system and event monitoring using Simple Network Management Protocol (SNMP), Remote Monitoring (RMON), Syslog, Cisco Discovery Protocol (CDP), and Network Time Protocol (NTP).
- Security Configuration— passwords, port security, physical security, and authentication using TACACS+.
- Configuration Checklist— summary of suggested configuration templates.

Basic Configuration

Features deployed with the majority of Catalyst networks are discussed in this section.

Catalyst Control Plane Protocols

This section introduces the protocols that run between switches under normal operation. A basic understanding of these protocols is helpful in tackling each section.

Supervisor Traffic

Most features enabled in a Catalyst network require two or more switches to cooperate, so there must be a controlled exchange of keepalive messages, configuration parameters, and management changes. Whether these protocols are Cisco proprietary, like CDP, or standards-based, like IEEE 802.1d (STP), all have certain elements in common when implemented on the Catalyst series.

In basic frame forwarding, user data frames originate from end systems, and their source address and destination address are not changed throughout Layer 2 (L2) switched domains. Content Addressable Memory (CAM) lookup-tables on each switch Supervisor Engine are populated by a source address learning process and indicate which egress port must forward each frame received. If the address learning process is incomplete (the destination is unknown or the frame is destined to a broadcast or multicast address), it is forwarded (flooded) out all ports in that VLAN.

The switch must also recognize which frames are to be switched through the system and which must be directed to the switch CPU itself (also known as the Network Management Processor [NMP]).

The Catalyst control plane is created using special entries in the CAM table called **system entries** in order to receive and direct traffic to the NMP on an internal switch port. Thus, by using protocols with well-known destination MAC addresses, control plane traffic can be separated from the data traffic. Issue **show CAM system** command on a switch to confirm this, as shown:

```
>show cam system

* = Static Entry. + = Permanent Entry. # = System Entry. R = Router Entry.
X = Port Security Entry
VLAN  Dest MAC/Route Des    [CoS]  Destination Ports or VCs / [Protocol Type]
----  ------------------    -----  ------------------------------------------
1     00-d0-ff-88-cb-ff  #          1/3
!--- NMP internal port. 1 01-00-0c-cc-cc-cc # 1/3 !--- CDP and so on. 1 01-00-0c-cc-
cc-cd # 1/3 !--- Cisco STP. 1 01-80-c2-00-00-00 # 1/3 !--- IEEE STP. 1 01-80-c2-00-
00-01 # 1/3 !--- IEEE flow control. 1 00-03-6b-51-e1-82 R# 15/1 !--- Multilayer
```

Switch Feature Card (MSFC) router. ...

Cisco has a reserved range of Ethernet MAC and protocol addresses, as shown. Each one is covered later in this document. However, a summary is presented in this table for convenience.

Feature	SNAP HDLC Protocol Type	Destination Multicast MAC
Port Aggregation Protocol (PAgP)	0x0104	01-00-0c-cc-cc-cc
Spanning Tree PVSTP+	0x010b	01-00-0c-cc-cc-cd
VLAN Bridge	0x010c	01-00-0c-cd-cd-ce
Unidirectional Link Detection (UDLD)	0x0111	01-00-0c-cc-cc-cc
Cisco Discovery Protocol	0x2000	01-00-0c-cc-cc-cc
Dynamic Trunking (DTP)	0x2004	01-00-0c-cc-cc-cc
STP Uplink Fast	0x200a	01-00-0c-cd-cd-cd
IEEE Spanning Tree 802.1d	N/A - DSAP 42 SSAP 42	01-80-c2-00-00-00
Inter Switch Link (ISL)	N/A	01-00-0c-00-00-00
VLAN Trunking (VTP)	0x2003	01-00-0c-cc-cc-cc
IEEE Pause, 802.3x	N/A - DSAP 81 SSAP 80	01-80-C2-00-00-00>0F

The majority of Cisco control protocols use an IEEE 802.3 SNAP encapsulation, including **LLC 0xAAAA03, OUI 0x00000C**, which can be seen on a LAN analyzer trace. Other common properties of these protocols include:

- These protocols assume point-to-point connectivity. Note that the deliberate use of multicast destination addresses enables two Catalysts to transparently communicate over non-Cisco switches, as devices that do not understand and intercept the frames simply flood them. However, point-to-multipoint connections through multi-vendor environments can result in inconsistent behavior and must generally be avoided.
- These protocols terminate at Layer 3 (L3) routers; they function only within a switch domain.
- These protocols receive prioritization over user data by ingress application-specific integrated circuit (ASIC) processing and scheduling.

After the introduction of the control protocol destination addresses, the source address must also

be described for completeness. Switch protocols use a MAC address taken from a bank of available addresses provided by an EPROM on the chassis. Issue the **show module** command in order to display the address ranges available to each module when it sources traffic such as STP bridge protocol data units (BPDUs) or ISL frames.

```
>show module

...
Mod MAC-Address(es)                        Hw     Fw          Sw
--- -------------------------------------- ------ ----------- ------------------
1   00-01-c9-da-0c-1e to 00-01-c9-da-0c-1f 2.2    6.1(3)      6.1(1d)
    00-01-c9-da-0c-1c to 00-01-c9-da-0c-1
    00-d0-ff-88-c8-00 to 00-d0-ff-88-cb-ff
!--- MACs for sourcing traffic. ... VLAN 1
```

VLAN 1

VLAN 1 has a special significance in Catalyst networks.

The Catalyst Supervisor Engine always uses the default VLAN, VLAN 1, to tag a number of control and management protocols when trunking, such as CDP, VTP and PAgP. All ports, including the internal sc0 interface, are configured by default to be members of VLAN 1. All trunks carry VLAN 1 by default, and in CatOS software versions earlier than 5.4, it was not possible to block user data in VLAN 1.

These definitions are needed in order to help clarify some well-used terms in Catalyst networking:

- The management VLAN is where sc0 resides; this VLAN can be changed.
- The native VLAN is defined as the VLAN to which a port returns when not trunking, and is the untagged VLAN on an 802.1Q trunk. By default, VLAN 1 is the native VLAN.
- In order to change the native VLAN, issue the **set vlan** *vlan-id mod/port* command.**Note:** Create the VLAN before you set it as the native VLAN of the trunk.

These are several good reasons to tune a network and alter the behavior of ports in VLAN 1:

- When the diameter of VLAN 1, like any other VLAN, gets large enough to be a risk to stability (particularly from an STP perspective) it needs to be pruned back. This is discussed in more detail in the In-Band Management section of this document.
- Control plane data on VLAN 1 must be kept separate from the user data in order to simplify troubleshooting and maximize available CPU cycles.
- L2 loops in VLAN 1 must be avoided when multilayer-campus networks are designed without STP, and trunking is still required to the access layer if there are multiple VLANs and IP subnets. To do this, manually clear VLAN 1 from trunk ports.

In summary, note this information about trunks:

- **CDP, VTP, and PAgP** updates are always forwarded on trunks with a VLAN 1 tag. This is the case even if VLAN 1 is cleared from the trunks and is not the native VLAN. If VLAN 1 is cleared for user data, these is no impact on control plane traffic that is still sent using VLAN 1.
- On an ISL trunk, DTP packets are sent on VLAN1. This is the case even if VLAN 1 is cleared from the trunk and is no longer the native VLAN. On an 802.1Q trunk, DTP packets are sent on the native VLAN. This is the case even if the native VLAN is cleared from the trunk.

- In PVST+, the **802.1Q IEEE BPDUs** are forwarded untagged on the common Spanning Tree VLAN 1 for interoperability with other vendors, unless VLAN 1 is cleared from the trunk. This is the case regardless of the native VLAN configuration. **Cisco PVST+ BPDUs** are sent and tagged for all other VLANs. Refer to the Spanning Tree Protocol section in this document for more details.
- 802.1s Multiple Spanning Tree (MST) BPDUs are always sent on VLAN 1 on both ISL and 802.1Q trunks. This applies even when VLAN 1 is cleared from the trunks.
- Do not clear or disable VLAN 1 on trunks between MST bridges and PVST+ bridges. But, in the case that VLAN 1 is disabled, the MST bridge must become root in order for all VLANs to avoid the MST bridge putting its boundary ports in the root-inconsistent state. Refer to Understanding Multiple Spanning Tree Protocol (802.1s) for details.

Recommendations

In order to keep a VLAN in an **up/up** state with no clients or hosts connected in that VLAN , you need to have at least one physical device connected in that VLAN. Otherwise, the VLAN has an **up/down** state. Currently, there is no command to put a VLAN interface **up/up** when there are no active ports in the switch for that VLAN.

If you do not want to connect a device, connect a loopback plug in any port for that VLAN. As an alternative, try a crossover cable that connects two ports in that VLAN on the same switch. This method forces the port up. Refer to the Loopback Plug section of Loopback Tests for T1/56K Lines for more information.

When a network is multihomed to service providers, the network acts as a transit network between two service providers. If the VLAN number received in a packet needs to be translated or changed when passed from one service provider to another service provider, it is advisable to use the QinQ feature in order to translate the VLAN number.

VLAN Trunking Protocol

Before you create VLANs, determine the VTP mode to be used in the network. VTP enables VLAN configuration changes to be made centrally on one or more switches. Those changes automatically propagate to all other switches in the domain.

Operational Overview

VTP is a L2 messaging protocol that maintains VLAN configuration consistency. VTP manages the addition, deletion, and renaming of VLANs on a network-wide basis. VTP minimizes misconfigurations and configuration inconsistencies that can cause a number of problems, such as duplicate VLAN names, incorrect VLAN-type specifications, and security violations. The VLAN database is a binary file and is stored in NVRAM on VTP servers separately from the configuration file.

The VTP protocol communicates between switches using an Ethernet destination multicast MAC address (**01-00-0c-cc-cc-cc**) and SNAP HDLC protocol type Ox2003. It does not work over non-trunk ports (VTP is a payload of ISL or 802.1Q), so messages cannot be sent until DTP has brought the trunk online.

Message types include summary advertisements every five minutes, subset advertisements and

request advertisements when there are changes, and joins when VTP pruning is enabled. The VTP configuration revision number is incremented by one with every change on a server, which then propagates the new table across the domain.

If a VLAN is deleted, ports that were once a member of that VLAN are placed in an inactive state. Similarly, if a switch in client mode is unable to receive the VTP VLAN table at boot-up (either from a VTP server or another VTP client), all ports in VLANs other than the default VLAN 1 are deactivated.

This table provides a feature comparison summary for various VTP modes:

Feature	**Server**	**Client**	**Transparent**	**Off**[1]
Source VTP messages	Yes	Yes	No	No
Listen to VTP messages	Yes	Yes	No	No
Forward VTP messages	Yes	Yes	Yes	No
Create VLANs	Yes	No	Yes (locally significant only)	Yes (locally significant only)
Remember VLANs	Yes	No	Yes (locally significant only)	Yes (locally significant only)

In VTP `transparent` mode, VTP updates are ignored (the VTP multicast MAC address is removed from the system CAM that is normally used to pick up control frames and direct them to the supervisor engine). As the protocol uses a multicast address, a switch in transparent mode (or another vendor switch) simply floods the frame to other Cisco switches in the domain.

[1] CatOS software release 7.1 introduces the option to disable VTP with use of the `off` mode. In VTP `off` mode, the switch behaves in a way that is very similar to the VTP `transparent` mode, except that `off` mode also suppresses the forwarding of VTP updates.

This table provides a summary of the initial configuration:

Feature	**Default Value**
VTP Domain Name	Null
VTP mode	Server
VTP version	Version 1 is enabled
VTP password	None
VTP Pruning	Disabled

VTP version 2 (VTPv2) includes this functional flexibility. However, it is not interoperable with VTP version 1 (VTPv1):

- Token Ring support
- Unrecognized VTP information support; switches now propagate values they cannot parse.
- Version-dependent transparent mode; `transparent` mode no longer checks domain name. This enables support of more than one domain across a transparent domain.
- Version number propagation; if VTPv2 is possible on all switches, all can be enabled through the configuration of a single switch.

Refer to Understanding and Configuring VLAN Trunk Protocol (VTP) for more information.

VTP Version 3

CatOS software release 8.1 introduces support for VTP version 3 (VTPv3). VTPv3 provides enhancements over the existing versions. These enhancements allow for:

- Support for extended VLANs
- Support for the creation and advertisement of private VLANs
- Support for VLAN instances and MST mapping propagation instances (which are supported in CatOS release 8.3)
- Improved server authentication
- Protection from accidental insertion of the "wrong" database into a VTP domain
- Interaction with VTPv1 and VTPv2
- The ability to be configured on a per-port basis

One of the major differences between VTPv3 implementation and the earlier version is the introduction of a VTP primary server. Ideally, there must be only one primary server in a VTPv3 domain, if the domain is not partitioned. Any changes that you make to the VTP domain must be executed on the VTP primary server in order to be propagated to the VTP domain. There can be multiple servers within a VTPv3 domain, which are also known as secondary servers. When a switch is configured to be a server, the switch becomes a secondary server by default. The secondary server can store the configuration of the domain but cannot modify the configuration. A secondary server can become the primary server with a successful takeover from the switch.

Switches that run VTPv3 only accept a VTP database with a higher revision number than the current primary server. This process differs significantly from VTPv1 and VTPv2, in which a switch always accepts a superior configuration from a neighbor in the same domain. This change with VTPv3 provides protection. A new switch that is introduced into the network with a higher VTP revision number cannot overwrite the VLAN configuration of the entire domain.

The VTPv3 also introduces an enhancement to how the VTP handles passwords. If you use the hidden password configuration option in order to configure a password as "hidden", these items occur:

- The password does not appear in plain text in the configuration. The secret hexadecimal format of the password is saved in the configuration.
- If you try to configure the switch as a primary server, you are prompted for the password. If your password matches the secret password, the switch becomes a primary server, which allows you to configure the domain.

Note: It is important to note that the primary server is only necessary when you need to modify the VTP configuration for any instance. A VTP domain can operate with no active primary server because the secondary servers ensure persistence of the configuration over reloads. The primary server state is exited for these reasons:

- A switch reload
- A high-availability switchover between the active and redundant supervisor engines
- A takeover from another server
- A change in the mode configuration
- Any VTP domain configuration change, such as a change in:VersionDomain nameDomain password

VTPv3 also allows the switches to participate in multiple instances of VTP. In this case, the same switch can be the VTP server for one instance and a client for another instance because the VTP modes are specific to different VTP instances. For example, a switch can operate in `transparent` mode for an MST instance while the switch is configured in `server` mode for a VLAN instance.

In terms of interaction with VTPv1 and VTPv2, the default behavior in all versions of VTP has been that the earlier versions of VTP simply drop the new version updates. Unless the VTPv1 and VTPv2 switches are in `transparent` mode, all VTPv3 updates are dropped. On the other hand, after VTPv3 switches receive a legacy VTPv1 or VTPv2 frame on a trunk, the switches pass a scaled-down version of their database update to the VTPv1 and VTPv2 switches. However, this information exchange is unidirectional in that no updates from VTPv1 and VTPv2 switches are accepted by the VTPv3 switches. On trunk connections, VTPv3 switches continue to send out scaled-down updates as well as full-fledged VTPv3 updates in order to cater to the existence of VTPv2 and VTPv3 neighbors across the trunk ports.

In order to provide VTPv3 support for extended VLANs, the format of the VLAN database, in which the VTP assigns 70 bytes per VLAN, is changed. The change allows for the coding of non-default values only, instead of the carrying of unmodified fields for the legacy protocols. Because of this change, 4K VLAN support is the size of the resulting VLAN database.

Recommendation

There is no specific recommendation on whether to use VTP `client/server` modes or VTP `transparent` mode. Some customers prefer the ease of management of VTP `client/server` mode despite some considerations noted later. The recommendation is to have two `server` mode switches in each domain for redundancy, typically the two distribution-layer switches. The rest of the switches in the domain must be set to `client` mode. When you implement `client/server` mode with the use of VTPv2, be mindful that a higher revision number is always accepted in the same VTP domain. If a switch that is configured in either VTP `client` or `server` mode is introduced into the VTP domain and has a higher revision number than the existing VTP servers, this overwrites the VLAN database within the VTP domain. If the configuration change is unintentional and VLANs are deleted, the overwrite can cause a major outage in the network. In order to ensure that `client` or `server` switches always have a configuration revision number that is lower than that of the server, change the client VTP domain name to something other than the standard name. Then revert back to the standard. This action sets the configuration revision on the client to 0.

There are pros and cons to the VTP ability to make changes easily on a network. Many enterprises prefer the cautious approach of VTP `transparent` mode for these reasons:

- It encourages good change control practice, as the requirement in order to modify a VLAN on a switch or trunk port has to be considered one switch at a time.
- It limits the risk of an administrator error that impacts the entire domain, such as the deletion of a VLAN by accident..
- There is no risk that a new switch introduced into the network with a higher VTP revision number can overwrite the entire domain VLAN configuration.

- It encourages VLANs to be pruned from trunks running to switches that do not have ports in that VLAN. This makes frame flooding more bandwidth-efficient. Manual pruning is also beneficial because it reduces the spanning tree diameter (see the DTP section of this document). Before pruning unused VLANs on port channel trunks, ensure that any ports connected to IP phones are configured as access ports with voice VLAN.
- The extended VLAN range in CatOS 6.x and CatOS 7.x, numbers 1025 through 4094, can only be configured in this way. For more information, see the Extended VLAN and MAC Address Reduction section of this document.
- VTP `transparent` mode is supported in Campus Manager 3.1, part of Cisco Works 2000. The old restriction that required at least one server in a VTP domain has been removed.

Sample VTP Commands	**Comments**
set vtp domain *name* password *x*	CDP checks names in order to help check for miscabling between domains. A simple password is a helpful precaution against unintentional changes. Beware of case-sensitive names or spaces if pasting.
set vtp mode transparent	
set vlan *vlan number* name *name*	Per switch that has ports in the VLAN.
set trunk *mod/port vlan range*	Enables trunks to carry VLANs where needed - default is all VLANs.
clear trunk *mod/port vlan range*	Limits STP diameter by manual pruning, such as on trunks from distribution layer to access layer, where the VLAN does not exist.

Note: Specifying VLANs with the **set** command only adds VLANs, and does not clear them. For example, the **set trunk x/y 1-10** command does not set the allowed list to just VLANs 1-10. Issue the **clear trunk x/y 11-1005** command in order to achieve the desired result.

Although token ring switching is outside the scope of this document, note that VTP `transparent` mode is not recommended for TR-ISL networks. The basis for token ring switching is that the whole domain forms a single distributed multi-port bridge, so every switch must have the same

VLAN information.

Other Options

VTPv2 is a requirement in token ring environments, where `client/server` mode is highly recommended.

VTPv3 provides the ability to implement tighter authentication and configuration revision control. VTPv3 essentially provides the same level of functionality, but with more enhanced security, as VTPv1/VTPv2 `transparent` mode offers. In addition, VTPv3 is partially compatible with the legacy VTP versions.

The benefits of pruning VLANs to reduce unnecessary frame flooding are advocated in this document. The **set vtp pruning enable** command prunes VLANs automatically, which stops the inefficient flooding of frames where they are not needed. Unlike manual VLAN pruning, automatic pruning does not limit the Spanning Tree diameter.

From CatOS 5.1, the Catalyst switches can map 802.1Q VLAN numbers greater than 1000 to ISL VLAN numbers. In CatOS 6.x, Catalyst 6500/6000 switches support 4096 VLANs in accordance with the IEEE 802.1Q standard. These VLANs are organized into these three ranges, only some of which are propagated to other switches in the network with VTP:

- normal-range VLANs: 1–1001
- extended-range VLANs: 1025–4094 (can only be propagated by VTPv3)
- reserved-range VLANs: 0, 1002—1024, 4095

The IEEE has produced a standards-based architecture in order to accomplish similar results as VTP. As a member of the 802.1Q Generic Attribute Registration Protocol (GARP), the Generic VLAN Registration Protocol (GVRP) allows VLAN management interoperability between vendors, but is outside the scope of this document.

Note: CatOS 7.x introduces the option to set VTP to `off` mode, a mode very similar to `transparent`. However, the switch does not forward VTP frames. This can be useful in some designs when trunking to switches outside of your administrative control.

Extended VLAN and MAC Address Reduction

The MAC address reduction feature enables extended-range VLAN identification. The enablement of MAC address reduction disables the pool of MAC addresses that are used for the VLAN spanning tree and leaves a single MAC address. This MAC address identifies the switch. CatOS software release 6.1(1) introduces MAC address reduction support for Catalyst 6500/6000 and Catalyst 4500/4000 switches to support 4096 VLANs in compliance with the IEEE 802.1Q standard.

Operation Overview

Switch protocols use a MAC address that is taken from a bank of available addresses that an EPROM on the chassis provides as part of the bridge identifiers for VLANs that run under PVST+. Catalyst 6500/6000 and Catalyst 4500/4000 switches support either 1024 or 64 MAC addresses, which depends on the chassis type.

Catalyst switches with 1024 MAC addresses do not enable MAC address reduction by default.

MAC addresses are allocated sequentially. The first MAC address in the range is assigned to VLAN 1. The second MAC address in the range is assigned to VLAN 2, and so on. This enables the switches to support 1024 VLANs with each VLAN using a unique bridge identifier.

Chassis Type	Chassis Address
WS-C4003-S1, WS-C4006-S2	1024
WS-C4503, WS-C4506	64[1]
WS-C6509-E,WS-C6509, WS-C6509-NEB, WS-C6506-E, WS-C6506, WS-C6009, WS-C6006, OSR-7609-AC, OSR-7609-DC	1024
WS-C6513, WS-C6509-NEB-A, WS-C6504-E, WS-C6503-E, WS-C6503, CISCO7603, CISCO7606, CISCO7609, CISCO7613	64[1]

[1] MAC address reduction is enabled by default for switches that have 64 MAC addresses, and the feature cannot be disabled.

For Catalyst series switches with 1024 MAC addresses, an enablement of MAC address reduction allows support of 4096 VLANs that run under PVST+ or 16 Multiple Instance STP (MISTP) instances to have unique identifiers without an increase in the number of MAC addresses that are required on the switch. MAC address reduction reduces the number of MAC addresses that are required by the STP from one per VLAN or MISTP instance to one per switch.

This figure shows that the bridge identifier MAC address reduction is not enabled. The bridge identifier consists of a 2-byte bridge priority and a 6-byte MAC address:

MAC address reduction modifies the STP bridge identifier portion of the BPDU. The original 2-byte priority field is split into two fields. This split results in a 4-bit bridge priority field and a 12-bit system ID extension that allows for VLAN numbering of 0 through 4095.

When you have MAC address reduction enabled on Catalyst switches in order to leverage extended range VLANs, enable MAC address reduction on all switches within the same STP domain. This step is necessary in order to keep the STP root calculations on all switches consistent. After you enable MAC address reduction, the root bridge priority becomes a multiple of 4096 plus the VLAN ID. The switches without MAC address reduction can claim root inadvertently because these switches have a finer granularity in the selection of the bridge ID.

Configuration Guidelines

You must follow certain guidelines when you configure extended VLAN range. The switch can allocate a block of VLANs from the extended range for internal purposes. For example, the switch can allocate the VLANs for the routed ports or Flex WAN modules. Allocation of the block of VLANs always starts from VLAN 1006 and goes up. If you have any VLANs within the range that the Flex WAN module requires, all the required VLANs are not allocated because the VLANs are never allocated from the user VLAN area. Issue the **show vlan** command or the **show vlan summary** command on a switch in order to display both the user-assigned and internal VLANs.

```
>show vlan summary

Current Internal Vlan Allocation Policy - Ascending

Vlan status    Count  Vlans
-------------  -----  -----------------------------------------
VTP Active         7   1,17,174,1002-1005

Internal           7   1006-1011,1016
!--- These are internal VLANs. >show vlan
---- -------------------------------- --------- ------- --------

1    default                          active    7       4/1-48


!--- Output suppressed. 1006 Online Diagnostic Vlan1 active 0 internal 1007 Online
Diagnostic Vlan2 active 0 internal 1008 Online Diagnostic Vlan3 active 0 internal
1009 Voice Internal Vlan active 0 internal 1010 Dtp Vlan active 0 internal 1011
Private Vlan Internal Vlan suspend 0 internal 1016 Online SP-RP Ping Vlan active 0
internal !--- These are internal VLANs.
```

Additionally, before you use the extended-range VLANs, you must delete any existing 802.1Q-to-ISL mappings. Also, in versions earlier than VTPv3, you must statically configure the extended VLAN on each switch with the use of VTP `transparent` mode. Refer to the *Extended-Range VLAN Configuration Guidelines* section of Configuring VLANs for more information.

Note: In software that is earlier than software release 8.1(1), you cannot configure the VLAN name for extended-range VLANs. This capability is independent of any VTP version or mode.

Recommendation

Try to maintain a consistent MAC address reduction configuration within the same STP domain. However, the enforcement of MAC address reduction on all network devices can be impractical when new chassis with 64 MAC addresses are introduced to the STP domain. MAC address reduction is enabled by default for switches that have 64 MAC addresses, and the feature cannot be disabled. Understand that, when two systems are configured with the same spanning-tree priority, the system without MAC address reduction has a better spanning-tree priority. Issue this command in order to enable or disable MAC address reduction:

```
set spantree macreduction enable | disable
```

Allocation of the internal VLANs is in ascending order and starts at VLAN 1006. Assign the user VLANs as close to VLAN 4094 as possible in order to avoid conflicts between the user VLANs and the internal VLANs. With Catalyst 6500 switches that run Cisco IOS® system software, you can configure the internal VLAN allocation in descending order. The Command-Line Interface (CLI) equivalent for CatOS software is not officially supported.

Autonegotiation

Ethernet/Fast Ethernet

Autonegotiation is an optional function of the IEEE Fast Ethernet (FE) standard (802.3u) that

enables devices to automatically exchange information over a link about **speed** and **duplex** abilities. Autonegotiation operates at Layer 1 (L1), and targets access layer ports where **transient users** such as PCs connect to the network.

Operational Overview

The most common cause of performance issues on 10/100 Mbps Ethernet links occur when one port on the link operates at half-duplex while the other is at full-duplex. This occasionally happens when one or both ports on a link are reset and the autonegotiation process does not cause both link partners to have the same configuration. It also happens when administrators reconfigure one side of a link and forget to reconfigure the other side. The typical symptoms of this are increasing frame check sequence (FCS), cyclic redundancy check (CRC), alignment, or runt counters on the switch.

Autonegotiation is discussed in detail in these documents. These documents include explanations of how autonegotiation works and configuration options.

- Configuring and Troubleshooting Ethernet 10/100Mb Half/Full Duplex Auto-Negotiation
- Troubleshooting Cisco Catalyst Switches to NIC Compatibility Issues

A common misconception about autonegotiation is that it is possible to manually configure one link partner for 100 Mbps full-duplex and autonegotiate to full-duplex with the other link partner. In fact, an attempt to do this results in a duplex mismatch. This is a consequence of one link partner autonegotiating, not seeing any autonegotiation parameters from the other link partner, and defaulting to half-duplex.

Most Catalyst Ethernet modules support 10/100 Mbps and half/full-duplex, but the **show port capabilities *mod/port*** command confirms this.

FEFI

Far end fault indication (FEFI) protects 100BASE-FX (fiber) and Gigabit interfaces, while autonegotiation protects 100BASE-TX (copper) against physical-layer/signaling related faults.

A **far end fault** is an error in the link that one station can detect while the other cannot, such as a disconnected TX-wire. In this example, the sending station could still receive valid data and detect that the link is good through the link-integrity-monitor. It does not detect that its transmission is not being received by the other station. A 100BASE-FX station that detects such a remote fault can modify its transmitted IDLE stream to send a special bit-pattern (referred to as the FEFI IDLE pattern) to inform the neighbor of the remote fault; the FEFI-IDLE pattern subsequently triggers a shutdown of the remote port (errdisable). Refer to the UDLD section of this document for more information on fault protection.

FEFI is supported by this hardware and these modules:

- Catalyst 5500/5000: WS-X5201R, WS-X5305, WS-X5236, WS-X5237, WS-U5538, and WS-U5539
- Catalyst 6500/6000 and 4500/4000: All 100BASE-FX modules and GE modules

Recommendation

Whether to configure autonegotiation on 10/100 links or to hard code speed and duplex ultimately

depends on the type of link partner or end device you have connected to a Catalyst switch port. Autonegotiation between end devices and Catalyst switches generally works well, and Catalyst switches are compliant with the IEEE 802.3u specification. However, problems can result when NIC or vendor switches do not conform exactly. Hardware incompatibility and other issues can also exist as a result of vendor-specific advanced features, such as auto-polarity or cabling integrity, that are not described in the IEEE 802.3u specification for 10/100 Mbps autonegotiation. Refer to Field Notice: Performance Issue with Intel Pro/1000T NICs connecting to CAT4K/6K for an example of this.

Anticipate that there will be some situations that require host, port speed, and duplex to be set. In general, follow these basic troubleshooting steps:

- Make sure that either autonegotiation is configured on both sides of the link or hard coding is configured on both sides.
- Check the CatOS release notes for common caveats.
- Verify the version of NIC driver or operating system you are running, as the latest driver or patch is often required.

As a rule, try to use autonegotiation first for any type of link partner. There are obvious benefits to configuring autonegotiation for transient devices like laptops. Ideally, autonegotiation also works well with non-transient devices such as servers and fixed workstations or from switch-to-switch and switch-to-router. For some of the reasons mentioned, negotiation issues can arise. In these cases, follow the basic troubleshooting steps outlined in the TAC links provided.

If the port speed is set to `auto` on a 10/100 Mbps Ethernet port, both speed and duplex are autonegotiated. Issue this command in order to set the port to auto:

```
set port speed port range auto
!--- This is the default.
```

If hard coding the port, issue these configuration commands:

```
set port speed port range 10 | 100 set port duplex port range full | half
```

In CatOS 8.3 and later, Cisco has introduced the optional **auto-10-100** keyword. Use the **auto-10-100** keyword on ports that support speeds of 10/100/1000 Mbps but where autonegotiation to 1000 Mbps is undesirable. Use of the **auto-10-100** keyword makes the port behave in the same way as a 10/100-Mbps port that has the speed set to **auto**. The speed and duplex are negotiated for 10/100-Mbps ports only, and the 1000-Mbps speed does not take part in the negotiation.

```
set port speed port_range auto-10-100
```

Other Options

When no autonegotiation is used between switches, L1 fault indication can also be lost for certain problems. It is helpful to use L2 protocols to augment failure detection, such as aggressive UDLD.

Gigabit Ethernet

Gigabit Ethernet (GE) has an autonegotiation procedure (IEEE 802.3z) that is more extensive than that for 10/100 Mbps Ethernet and is used to exchange flow-control parameters, remote fault information, and duplex information (even though Catalyst series GE ports only support full-duplex mode).

Note: 802.3z has been superseded by IEEE 802.3:2000 specs. Refer to IEEE Standards On Line LAN/MAN Standards Subscription: Archives for more information.

Operational Overview

GE port negotiation is enabled by default, and the ports on both ends of a GE link must have the same setting. Unlike FE, the GE link does not come up if the autonegotiation setting differs on the ports at each end of the link. However, the only condition that is required for an autonegotiation-disabled port to link up is a valid Gigabit signal from the far end. This behavior is independent of the autonegotiation configuration of the far end. For example, assume that there are two devices, A and B. Each device can have autonegotiation enabled or disabled. This table is a list of possible configurations and respective link states:

Negotiation	B Enabled	B Disabled
A Enabled	`up` on both sides	A `down`, B `up`
A Disabled	A `up`, B `down`	`up` on both sides

In GE, synchronization and autonegotiation (if they are enabled) are performed upon link startup through the use of a special sequence of reserved link code words.

Note: There is a dictionary of valid words and not all possible words are valid in GE.

The life of a GE connection can be characterized in this way:

A loss of synchronization means that the MAC detects a link down. Loss of synchronization applies whether autonegotiation is enabled or disabled. Synchronization is lost under certain failed conditions, such as the receipt of three invalid words in succession. If this condition persists for 10 ms, a "sync fail" condition is asserted and the link is changed to the `link_down` state. After synchronization is lost, another three consecutive valid idles are necessary in order to resynchronize. Other catastrophic events, such as a loss of receive (Rx) signal, causes a link-down event.

Autonegotiation is a part of the linkup process. When the link is up, autonegotiation is over. However, the switch still monitors the status of the link. If autonegotiation is disabled on a port, the "autoneg" phase is no longer an option.

The GE copper specification (1000BASE-T) does support autonegotiation through a Next Page Exchange. Next Page Exchange allows autonegotiation for 10/100/1000-Mbps speeds on copper ports.

Note: The GE fiber specification only makes provisions for the negotiation of duplex, flow control, and remote fault detection. GE fiber ports do not negotiate port speed. Refer to sections 28 and 37 of the IEEE 802.3-2002 specification for more information on autonegotiation.

Synchronization restart delay is a software feature that controls the total autonegotiation time. If autonegotiation is not successful within this time, the firmware restarts autonegotiation in case there is a deadlock. The **set port sync-restart-delay** command only has an effect when autonegotiation is set to `enable`.

Recommendation

Enabling autonegotiation is much more critical in a GE environment than in a 10/100 environment. In fact, autonegotiation must only be disabled on switch ports that attach to devices not capable of supporting negotiation or where connectivity issues arise from interoperability issues. Cisco recommends that Gigabit negotiation be enabled (default) on all switch-to-switch links and generally all GE devices. Issue this command in order to enable autonegotiation:

```
set port negotiation port range enable
!--- This is the default.
```

One known exception is when there is a connection to a Gigabit Switch Router (GSR) running Cisco IOS Software earlier than release 12.0(10)S, the release that added flow control and autonegotiation. In this case, turn off those two features, or the switch port reports `not connected`, and the GSR reports errors. This is a sample command sequence:

```
set port flowcontrol receive port range off set port flowcontrol send port range off
set port negotiation port range disable
```

Switch-to-server connections must be looked at on a case-by-case basis. Cisco customers have encountered issues with Gigabit negotiation on Sun, HP, and IBM servers.

Other Options

Flow control is an optional part of the 802.3x specification and must be negotiated if used. Devices can or cannot be capable of sending and/or responding to a `PAUSE` frame (**well known MAC 01-80-C2-00-00-00 0F**). Also, they can not agree to the flow-control request of the far-end neighbor. A port with an input buffer that is filling up sends a `PAUSE` frame to its link partner, which stops the transmission, and holds any additional frames in the link partner output buffers. This does not solve any steady-state over-subscription problem, but effectively makes the input buffer larger by some fraction of the partner output buffer during bursts.

This feature is best used on links between access-ports and end hosts, where the host output buffer is potentially as large as their virtual memory. Switch-to-switch use has limited benefits.

Issue these commands in order to control this on the switch ports:

```
set port flowcontrol mod/port receive | send off |on | desired

>show port flowcontrol
```

```
 Port   Send FlowControl    Receive FlowControl   RxPause TxPause
        admin    oper       admin    oper
------  -------- --------   -------- --------     ------- -------
 6/1    off      off        on       on           0       0
 6/2    off      off        on       on           0       0
 6/3    off      off        on       on           0       0
```

Note: All Catalyst modules respond to a `PAUSE` frame if negotiated. Some modules (for example, WS-X5410, WS-X4306) never send `PAUSE` frames even if they negotiate to do so, as they are non-blocking.

Dynamic Trunking Protocol

Encapsulation Type

Trunks extend VLANs between devices by temporarily identifying and tagging (link-local) the original Ethernet frames, thus they enable them to be multiplexed over a single link. This also ensures the separate VLAN broadcast and security domains are maintained between switches. CAM tables maintain the frame-to-VLAN mapping inside the switches.

Trunking is supported on several types of L2 media, including ATM LANE, FDDI 802.10, and Ethernet, although only the latter is be presented here.

ISL Operational Overview

Cisco proprietary identification or tagging scheme, ISL, has been in use for many years. The 802.1Q IEEE standard is also available.

By totally encapsulating the original frame in a two-level tagging scheme, ISL is effectively a tunneling protocol and has the additional benefit of carrying non-Ethernet frames. It adds a 26-byte header and 4-byte FCS to the standard Ethernet frame - the larger Ethernet frames are expected and handled by ports configured to be trunks. ISL supports 1024 VLANs.

ISL Frame Format

40 Bits	**4 Bits**	**4 Bits**	**48 Bits**	**16 Bits**	**24 Bits**	**24 Bits**	**15 Bits**	**Bit**	**16 Bits**	**16 Bits**	**Variable length**	**32 Bits**
Dest. Addr	Type	USER	SA	LEN	SNAP LLC	HSA	VLAN	BPDU	INDEX	Reserve	Encapsulated Frame	FCS
01-00-0c-00-00					AAAA03	000 00C						

Refer to InterSwitch Link and IEEE 802.1Q Frame Format for more information.

802.1Q Operational Overview

The IEEE 802.1Q standard specifies much more than encapsulation types, including Spanning Tree enhancements, GARP (see the VTP section of this document), and 802.1p Quality of Service (QoS) tagging.

The 802.1Q frame format preserves the original Ethernet source address and destination address, yet switches must now expect baby-giant frames to be received, even on access ports where hosts can use tagging in order to express 802.1p user priority for QoS signaling. The tag is 4 bytes, so 802.1Q Ethernet v2 frames are 1522 bytes, an IEEE 802.3ac working group achievement. 802.1Q also supports numbering space for 4096 VLANs.

All data frames transmitted and received are 802.1Q-tagged except for those on the native VLAN (there is an implicit tag based on the ingress switch port configuration). Frames on the native VLAN are always transmitted untagged and normally received untagged. However, they can also be received tagged.

Refer to VLAN Standardization via IEEE 802.10 and Get IEEE 802 for more details.

802.1Q/801.1p Frame Format

		Tag Header						
		TPID	**TCI**					
48 bits	48 bits	**16 bits**	**3 bits**	**1 bit**	**12 bits**	16 bits	Variable length	32 bits
DA	SA	**TPID**	**Priority**	**CFI**	**VLAN ID**	Length/ Type	Data with PAD	FCS
		0x8100	**0 - 7**	**0-1**	**0-4095**			

Recommendation

As all newer hardware supports 802.1Q (and some only supports 802.1Q, such as the Catalyst 4500/4000 series and CSS 11000), Cisco recommends that all new implementations follow the IEEE 802.1Q standard and older networks gradually migrate from ISL.

The IEEE standard allows vendor interoperability. This is advantageous in all Cisco environments as new host 802.1p capable NICs and devices become available. Although both ISL and 802.1Q implementations are mature, the IEEE standard will ultimately have greater field exposure and greater third party support, such as network analyzer support. The lower encapsulation overhead of 802.1Q compared to ISL is a minor point in favor of 802.1Q as well.

As the encapsulation type is negotiated between switches using DTP, with ISL chosen as the winner by default if both ends support it, it is necessary to issue this command in order to specify dot1q:

```
set trunk mod/port mode dot1q
```

If VLAN 1 is cleared from a trunk, as discussed in the In-Band Management section of this document, although no user data is transmitted or received, the NMP continues to pass control protocols such as CDP and VTP on VLAN 1.

Also, as discussed in the VLAN 1 section of this document, CDP, VTP, and PAgP packets are always sent on VLAN 1 when trunking. When using dot1q encapsulation, these control frames are tagged with VLAN 1 if the native VLAN of the switch is changed. If dot1q trunking to a router is enabled and the native VLAN is changed on the switch, a sub-interface in VLAN 1 is needed to receive the tagged CDP frames and provide CDP neighbor visibility on the router.

Note: There is a potential security consideration with dot1q caused by the implicit tagging of the native VLAN, as it can be possible to send frames from one VLAN to another without a router. Refer to Are there Vulnerabilities in VLAN Implementations? for further details. The workaround is to use a VLAN ID for the native VLAN of the trunk that is not used for end user access. The majority of Cisco customers leave VLAN 1 as the native VLAN on a trunk and assign access ports to VLANs other than VLAN 1 in order to achieve this simply.

Trunking Mode

DTP is the second generation of Dynamic ISL (DISL), and exists in order to ensure that the different parameters involved in sending ISL or 802.1Q frames, such as the configured encapsulation type, native VLAN, and hardware capability, are agreed upon by the switches at either end of a trunk. This also helps protect against non-trunk ports flooding tagged frames, a potentially serious security risk, by ensuring that ports and their neighbors are in consistent states.

Operational Overview

DTP is a L2 protocol that negotiates configuration parameters between a switch port and its neighbor. It uses another multicast MAC address (**01-00-0c-cc-cc-cc**) and a SNAP protocol type of 0x2004. This table is a summary of the configuration modes:

Mode	Function	DTP Frames Transmitted	Final State (Local Port)
`Auto (default)`	Makes the port willing to convert the link to a trunk. The port becomes a trunk port if the neighboring port is set to `on` or `desirable` mode.	Yes, periodic.	Trunking
`On`	Puts the port into permanent trunking mode and negotiates to convert the link into a trunk. The port becomes a trunk port	Yes, periodic.	Trunking, **unconditionally**.

	even if the neighboring port does not agree to the change.		
`Nonegotiate`	Puts the port into permanent trunking mode but prevents the port from generating DTP frames. You must configure the neighboring port manually as a trunk port to establish a trunk link. This is useful for devices that do not support DTP.	No	Trunking, **unconditionally**.
`Desirable`	Makes the port actively attempt to convert the link to a trunk link. The port becomes a trunk port if the neighboring port is set to `on`, `desirable`, or `auto` mode.	Yes, periodic.	It ends up in trunking state only if the remote mode is `on`, `auto`, or `desirable`.
`Off`	Puts the port into permanent non-trunking mode and negotiates to convert the link into a non-trunk link. The port becomes a non-trunk port even if the neighboring port does not agree to the change.	No in steady state, but transmits informs to speed up remote end detection after the change from `on`.	Non-trunking

These are some highlights of the protocol:

- DTP assumes a point-to-point connection, and Cisco devices only support 802.1Q trunk ports that are point-to-point.
- During DTP negotiation, the ports do not participate in STP. Only after the port becomes one of the three DTP types (access, ISL, or 802.1Q) does the port be added to STP. Otherwise PAgP, if configured, is the next process to run before the port participates in STP.
- If the port is trunking in ISL mode, DTP packets are sent out on VLAN 1, otherwise (for 802.1Q trunking or non-trunking ports) they are sent out on the native VLAN.
- In `desirable` mode, DTP packets transfer the**VTP domain name** (which must match for a negotiated trunk to come up), plus trunk configuration and **admin status**.
- Messages are sent every second during negotiation, and every 30 seconds after that.
- Be sure to understand that modes `on`, `nonegotiate`, and `off` explicitly specify in which state the port ends up. A bad configuration can lead to a dangerous/inconsistent state where one side is trunking and the other is not.

- A port in `on`, `auto`, or `desirable` mode sends DTP frames periodically. If a port in `auto` or `desirable` mode does not see a DTP packet in five minutes, it is set to non-trunk.

Refer to Configuring ISL Trunking on Catalyst 5500/5000 and 6500/6000 Family Switches for more ISL details. Refer to Trunking Between Catalyst 4500/4000, 5500/5000, and 6500/6000 Series Switches Using 802.1Q Encapsulation with Cisco CatOS System Software for more 802.1Q details.

Recommendation

Cisco recommends an explicit trunk configuration of `desirable` at both ends. In this mode, network operators can trust syslog and command line status messages that a port is up and trunking, unlike `on` mode, which can make a port appear up even though the neighbor is misconfigured. In addition, `desirable` mode trunk provides stability in situations where one side of the link cannot become a trunk or drops trunk state. Issue this command in order to set `desirable` mode:

```
set trunk mod/port desirable ISL | dot1q
```

Note: Set trunk to `off` on all non-trunk ports. This helps eliminate wasted negotiation time when bringing host ports up. This command is also executed when the **set port host** command is used; refer to the STP section for more information. Issue this command in order to disable a trunk on a range of ports:

```
set trunk port range off
!--- Ports are not trunking; part of the set port host command.
```

Other Options

Another common customer configuration uses `desirable` mode only at the distribution layer and the simplest default configuration (`auto` mode) at the access layer.

Some switches, such as a Catalyst 2900XL, Cisco IOS routers, or other vendor devices, do not currently support trunk negotiation through DTP. You can use `nonegotiate` mode on Catalyst 4500/4000, 5500/5000, and 6500/6000 switches in order to set a port to trunk unconditionally with these devices, which can help standardize on a common setting across the campus. Also, you can implement `nonegotiate` mode in order to reduce the "overall" link initialization time.

Note: Factors such as the channel mode and STP configuration can also affect the initialization time.

Issue this command in order to set `nonegotiate` mode:

```
set trunk mod/port nonegotiate ISL | dot1q
```

Cisco recommends `nonegotiate` when there ia a connection to a Cisco IOS router because when bridging is performed, some DTP frames received from **on** mode can get back into the trunk port. Upon reception of the DTP frame, the switch port tries to renegotiate (orbring the trunk down and up) unnecessarily. If `nonegotiate` is enabled, the switch does not send DTP frames.

Spanning Tree Protocol

Basic Considerations

Spanning Tree Protocol (STP) maintains a loop-free L2 environment in redundant switched and bridged networks. Without STP, frames loop and/or multiply indefinitely, which causes a network meltdown as all devices in the broadcast domain are interrupted continuously by high traffic.

Although in some respects STP is a mature protocol initially developed for slow software-based bridge specifications (IEEE 802.1d), it can be complex to implement well in large switched networks with many VLANs, many switches in a domain, multi-vendor support, and newer IEEE enhancements.

For future reference, CatOS 6.x continues to take on new STP development, such as MISTP, loop-guard, root-guards, and BPDU arrival time skew detection. In addition, further standardized protocols are available in CatOS 7.x, such as IEEE 802.1s shared Spanning Tree and IEEE 802.1w rapid convergence Spanning Tree.

Operational Overview

The root bridge election per VLAN is won by the switch with the lowest root Bridge Identifier (BID). The BID is the bridge priority combined with the switch MAC address.

Initially, BPDUs are sent from all switches, containing the BID of each switch and the path cost to reach that switch. This enables the root bridge and the lowest-cost path to the root to be determined. Additional configuration parameters carried in BPDUs from the root override those that are locally configured so that the whole network uses consistent timers.

The topology then converges through these steps:

1. A single root bridge is elected for the entire Spanning Tree domain.
2. One root port (facing the root bridge) is elected on every non-root bridge.
3. A designated port is elected for BPDU forwarding on every segment.
4. Non-designated ports become blocking.

Refer to Configuring Spanning Tree for more information.

Basic Timer Defaults (seconds)	**Name**	**Function**
	`Hello`	

2		Controls sending of BPDUs.
15	Forward Delay (Fwddelay)	Controls how long a port spends in listening and learning state and influences the topology change process (see next section).
20	Maxage	Controls how long the switch maintains the current topology before it looks for an alternative path. After the Maxage seconds, a BPDU is considered stale and the switch looks for a new root port from the pool of blocking ports. If no blocked port is available, it claims to be the root itself on the designated ports.

Port States	Meaning	Default timing to next state
Disabled	Administratively down.	N/A
Blocking	Receiving BPDUs and stopping user data.	Monitor reception of BPDUs. Wait 20 seconds for Maxage expiration or immediate change if direct/local link failure detected.
Listening	Sending or receiving BPDUs to check whether return to blocking needed.	Fwddelay timer (wait 15 seconds)
Learning	Building topology/CAM table.	Fwddelay timer (wait 15 seconds)
Forwarding	Sending/receiving data.	
	Total basic topology change:	**20 + 2 (15) = 50 seconds if waiting for Maxage to expire, or 30 seconds for direct link failure**

The two types of BPDUs in STP are configuration BPDUs and Topology Change Notification (TCN) BPDUs.

Configuration BPDU Flow

Configuration BPDUs are sourced every hello-interval from every port on the root bridge and subsequently flow to all leaf switches in order to maintain the state of the Spanning Tree. In steady state, the BPDU flow is unidirectional: root ports and blocking ports only receive configuration BPDUs, while designated ports only send configuration BPDUs.

For every BPDU received by a switch from the root, a new one is processed by the Catalyst

central NMP and sent out containing the root information. In other words, if the root bridge is lost or all paths to the root bridge are lost, BPDUs stop being received (until the maxage timer starts re-election).

TCN BPDU Flow

TCN BPDUs are sourced from leaf switches and flow towards the root bridge when a topology change is detected in the spanning tree. Root ports only send TCNs, and designated ports only receive TCNs.

The TCN BPDU travels toward the root ridge and is acknowledged at each step, so this is a reliable mechanism. Once it arrives at the root bridge, the root bridge alerts the entire domain that a change has occurred by sourcing Configuration BPDUs with the TCN flag set for **`maxage + fwddelay`** time (35 seconds by default). This causes all switches to change their normal CAM aging time from five minutes (by default) to the interval specified by **`fwddelay`** (15 seconds by default). Refer to Understanding Spanning Tree Protocol Topology Changes for more details.

Spanning Tree Modes

There are three different ways to correlate VLANs with Spanning Tree:

- A single Spanning Tree for all VLANs, or mono Spanning Tree Protocol, such as IEEE 802.1Q
- A Spanning Tree per VLAN, or shared Spanning Tree, such as Cisco PVST
- A Spanning Tree per set of VLANs, or multiple Spanning Tree, such as Cisco MISTP and IEEE 802.1s

A mono Spanning Tree for all VLANs allows only one active topology and therefore no load balancing. An STP blocked port blocks for all VLANs and carries no data.

One Spanning Tree per VLAN allows load balancing but requires more BPDU CPU processing as the number of VLANs increases. The CatOS release notes provide guidance on the number of logical ports recommended in the Spanning Tree per switch. For example, the Catalyst 6500/6000 Supervisor Engine 1 formula is as such:

number of ports + (number of trunks * number of VLANs on trunks) < 4000

Cisco MISTP and the new 802.1s standard allow the definition of only two active STP instances/topologies, and the mapping of all VLANs to either of these two trees. This technique allows STP to scale to many thousands of VLANs while load balancing is enabled.

BPDU Formats

In order to support the IEEE 802.1Q standard, the existing Cisco STP implementation was extended to become PVST+ by adding support for tunneling across an IEEE 802.1Q mono Spanning Tree region. PVST+ is therefore compatible with both IEEE 802.1Q MST and Cisco PVST protocols and does not require extra commands or configuration. In addition, PVST+ adds verification mechanisms in order to ensure that there is no configuration inconsistency of port trunking and VLAN IDs across switches.

These are some operational highlights of the PVST+ protocol:

- PVST+ interoperates with 802.1Q mono Spanning Tree through the so-called Common

Spanning Tree (CST) over an 802.1Q trunk. The CST is always on VLAN 1, so this VLAN needs to be enabled on the trunk to interoperate with other vendors. CST BPDUs are transmitted, always untagged, to the IEEE Standard Bridge-Group (MAC Address 01-80-c2-00-00-00, DSAP 42, SSAP 42). For completeness of description, a parallel set of BPDUs are also transmitted to the Cisco shared Spanning Tree MAC address for VLAN 1.

- PVST+ tunnels PVST BPDUs across 802.1Q VLAN regions as multicast data. Cisco shared Spanning Tree BPDUs are transmitted to MAC address 01-00-0c-cc-cc-cd (SNAP HDLC protocol type 0x010b) for each VLAN on a trunk. BPDUs are untagged on the native VLAN and tagged for all other VLANs.
- PVST+ checks port and VLAN inconsistencies. PVST+ blocks those ports that receive inconsistent BPDUs in order to prevent forwarding loops. It also notifies users through syslog messages about any configuration mismatch.
- PVST+ is backward-compatible with existing Cisco switches running PVST on ISL trunks. ISL-encapsulated BPDUs are still transmitted or received using the IEEE MAC address. In other words, each BPDU type is link-local; there are no translation issues.

Recommendation

All Catalyst switches have STP enabled by default. This is recommended even if a design is chosen that does not include L2 loops so that STP is not enabled in the sense that it is actively maintaining a blocked port.

```
set spantree enable all
!--- This is the default.
```

Cisco recommends that STP is left enabled for these reasons:

- If there is a loop (induced by mispatching, bad cable, and so on.), STP prevents detrimental effects to the network caused by multicast and broadcast data.
- Protection against an EtherChannel breaking down.
- Most networks are configured with STP, which gives it maximum field exposure. More exposure generally equates to stable code.
- Protection against dual attached NICs misbehaving (or bridging enabled on servers).
- The software for many protocols (such as PAgP, IGMP snooping, and trunking) is closely related to STP. Running without STP can lead to undesirable results.

Do not change timers, as this can adversely affect stability. The majority of networks deployed are not tuned. The simple STP timers accessible through the command line, such as hello-interval and Maxage, are themselves comprised of a complex set of other assumed and intrinsic timers, so it is difficult to tune timers and consider all the ramifications. Moreover, there is the danger of undermining UDLD protection.

Ideally, keep user traffic off the management VLAN. Especially with older Catalyst switch processors, it is best to avoid problems with STP by keeping the management VLAN separate from user data. One end station that misbehaves could potentially keep the supervisor engine processor so busy with broadcast packets that it can miss one or more BPDUs. However, newer switches with more powerful CPUs and throttling controls relieve this consideration.. See the In-Band Management section of this document for more details.

Do not over-design redundancy. This can lead to a troubleshooting nightmare - too many blocking ports adversely affect long-term stability. **Keep the total SPT diameter under seven hops**. Try to design to the Cisco multilayer model, with its smaller switched domains, STP triangles, and deterministic blocked ports (as explained in Gigabit Campus Network Design—Principles and Architecture) wherever possible.

Influence and know where Root functionality and blocked ports reside, and document them on the topology diagram. The blocked ports are where STP troubleshooting begins - what made them change from blocking to forwarding is often the key part of root cause analysis. **Choose the distribution and core layers as the location of root/secondary Root**, since these are considered the most stable parts of the network. Check for optimal L3 and HSRP overlay with L2 data-forwarding paths. This command is a macro that configures the bridge priority; root sets it much lower than the default (32768), while root secondary sets it reasonably lower than the default:

```
set spantree root secondary vlan range
```

Note: This macro sets the root priority to be either 8192 (by default), the current root priority minus 1 (if another root bridge is known), or the current root priority (if its MAC address is lower then the current root).

Prune unnecessary VLANs off trunk-ports (a bidirectional exercise). This limits the diameter of STP and NMP processing overhead on portions of the network where certain VLANs are not required. VTP automatic pruning does not remove STP from a trunk. Refer to the VTP section of this document for more information. The default VLAN 1 can also be removed from trunks using CatOS 5.4 and later.

Refer to Spanning Tree Protocol Problems and Related Design Considerations for additional information.

Other Options

Cisco has another STP known as**VLAN-bridge**. This protocol operates using a destination MAC address of **01-00-0c-cd-cd-ce** and protocol type of 0x010c.

This is most useful if there is a need to bridge non-routable or legacy protocols between VLANs without interfering with the IEEE Spanning Tree instance(s) running on those VLANs. If VLAN interfaces for non-bridged traffic become blocked for L2 traffic (and this could easily happen if they participated in the same STP as IP VLANs), the overlaying L3 traffic gets inadvertently pruned off as well - an unwanted side-effect. VLAN-bridge is therefore a separate instance of STP for bridged protocols, which provides a separate topology that can be manipulated without affecting IP traffic.

The Cisco recommendation is to run VLAN-bridge if bridging is required between VLANs on Cisco routers such as the MSFC.

PortFast

PortFast is used to bypass normal Spanning Tree operation on access ports to speed up connectivity between end-stations and the services they need to connect to after link initialization.

On some protocols, such as IPX/SPX, it is important to see the access port in forwarding mode immediately after the link state has gone up in order to avoid GNS problems.

Refer to Using Portfast and Other Commands to Fix Workstation Startup Connectivity Delays for more information.

Operational Overview

PortFast skips the normal `listening` and `learning` states of STP by moving a port directly from `blocking` to `forwarding` mode after the link is known to be running. If this feature is not enabled, STP discards all user data until it decides that the port is ready to be moved to `forwarding` mode. This could take up to twice the `ForwardDelay` time (a total of 30 seconds by default).

PortFast mode also prevents an STP TCN from being generated each time a port state changes from `learning` to `forwarding`. TCNs are not a problem by themselves, but if a wave of TCNs hit the root bridge (typically in the morning when people turn on their PCs), it could extend convergence time unnecessarily.

STP PortFast is particularly important in both multicast CGMP and Catalyst 5500/5000 MLS networks. TCNs in these environments can cause the static CGMP CAM table entries to be aged out, which results in multicast packet loss until the next IGMP report, and/or flush MLS cache entries that then need to be rebuilt and could result in a router CPU spike, depending on the size of the cache. (Catalyst 6500/6000 MLS implementations and multicast entries learned from IGMP snooping are not affected.)

Recommendation

Cisco recommends that STP PortFast be enabled for all active host ports and disabled for switch-switch links and ports not in use.

Trunking and channeling must also be disabled for all host ports. Each access port is enabled by default for trunking and channeling, yet switch neighbors are not expected by design on host ports. If these protocols are left to negotiate, the subsequent delay in port activation can lead to undesirable situations in which initial packets from workstations, such as DHCP requests, are not forwarded.

CatOS 5.2 introduced a macro command, **set port host *port range*** that implements this configuration for access ports and helps autonegotiation and connection performance significantly:

```
set port host port range
!--- Macro command for these commands: set spantree portfast port range enable set
trunk port range off set port channel port range mode off
```

Note: PortFast does not mean that Spanning Tree is not run at all on those ports. BPDUs are still sent, received, and processed.

Other Options

PortFast BPDU-guard provides a way to prevent loops by moving a non-trunking port into an

`errdisable` state when a BPDU is received on that port.

A BPDU packet must never be received on an access port configured for PortFast, since host ports must not be attached to switches. If a BPDU is observed, it indicates an invalid and possibly dangerous configuration that needs administrative action. When the BPDU-guard feature is enabled, Spanning Tree shuts down PortFast-configured interfaces that receive BPDUs instead of putting them into the STP `blocking` state.

The command works on a per-switch basis, not per-port, as shown:

```
set spantree portfast bpdu-guard enable
```

The network manager is notified by an SNMP trap or syslog message if the port goes down. It is also possible to configure an automatic recovery time for errdisabled ports. Refer to the UDLD section of this document for more details. For more information, refer to Spanning Tree Portfast BPDU Guard Enhancement.

Note: PortFast for trunk ports was introduced in CatOS 7.x and has no effect on trunk ports in earlier releases. PortFast for trunk ports is designed to increase convergence times for L3 networks. To complement this feature, CatOS 7.x also introduced the possibility of the configuration of PortFast BPDU-guard on a per-port basis.

UplinkFast

UplinkFast provides fast STP convergence after a direct link failure in the network access layer. It does not modify STP, and its purpose is to speed up convergence time in a specific circumstance to less than three seconds, rather than the typical 30-second delay. Refer to Understanding and Configuring the Cisco Uplink Fast Feature for more information.

Operational Overview

Using the Cisco multilayer design model at the access layer, if the forwarding uplink is lost, the blocking uplink is immediately moved to a `forwarding` state without waiting for `listening` and `learning` states.

An uplink group is a set of ports per VLAN that can be thought of as a root port and backup root port. Under normal conditions, the root port(s) are assuring connectivity from the access toward the root. If this primary root-connection fails for any reason, the backup root link kicks in immediately without having to go through typical 30 seconds of convergence delay.

Because this effectively bypasses the normal STP topology change-handling process (`listening` and `learning`), an alternate topology correction mechanism is needed in order to update switches in the domain that local end stations are reachable through an alternate path. The access layer switch running UplinkFast also generates frames for each MAC address in its CAM to a multicast MAC address (01-00-0c-cd-cd-cd, HDLC protocol 0x200a) to update the CAM table in all switches in the domain with the new topology.

Recommendation

Cisco recommends that UplinkFast be enabled for switches with blocked ports, typically at the access layer. Do not use on switches without the implied topology knowledge of a backup root link - typically distribution and core switches in the Cisco multilayer design. It can be added without disruption to a production network. Issue this command in order to enable UplinkFast:

```
set spantree uplinkfast enable
```

This command also sets the **bridge priority** high in order to minimize the risk of this becoming a root bridge and the **port priority** high to minimize becoming a designated port, which breaks the functionality. When you restore a switch that had UplinkFast enabled, the feature has to be disabled, the uplink database cleared with "clear uplink," and the bridge priorities restored manually.

Note: The **all protocols** keyword for the UplinkFast command is needed when the protocol filtering feature is enabled. As the CAM records the protocol type as well as MAC and VLAN information when protocol filtering is enabled, an UplinkFast frame needs to be generated for each protocol on each MAC address. The **rate** keyword indicates the packets per second of the uplinkfast topology update frames. The default is recommended. You do not need to configure BackboneFast with Rapid STP (RSTP) or IEEE 802.1w because the mechanism is natively included and automatically enabled in RSTP.

BackboneFast

BackboneFast provides rapid convergence from indirect link failures. With the added functionality to STP, convergence times can typically be reduced from the default of 50 seconds to 30 seconds.

Operational Overview

The mechanism is initiated when a root port or blocked port on a switch receives inferior BPDUs from its designated bridge. This can happen when a downstream switch has lost its connection to the root and starts to send its own BPDUs in order to elect a new root. An **inferior BPDU** identifies a switch as both the root bridge and the designated bridge.

Under normal Spanning Tree rules, the receiving switch ignores inferior BPDUs for the configured maximum aging time, 20 seconds by default. However, with BackboneFast, the switch sees the inferior BPDU as a signal that the topology could have changed, and tries to determine whether it has an alternate path to the root bridge using Root Link Query (RLQ) BPDUs. This protocol addition allows a switch to check whether the root is still available, moves a `blocked` port to `forwarding` in less time, and notifies the isolated switch that sent the inferior BPDU that the root is still there.

These are some highlights of the protocol operation:

- A switch transmits the RLQ packet out the root port only (that is, towards the root bridge).
- A switch that receives a RLQ can reply either if it is the root switch, or if it knows it has lost connection with the root. If it does not know these facts, it must forward the query out its root port.
- If a switch has lost connection to the root, it must reply in the negative to this query.

- The reply must be sent out only the port from which the query came.
- The root switch must always respond to this query with a positive reply.
- If the reply is received on a non-root port, it is discarded.

STP convergence times can therefore be reduced by up to 20 seconds, as maxage does not need to expire.

Refer to Understanding and Configuring Backbone Fast on Catalyst Switches for more information.

Recommendation

The Cisco recommendation is to enable BackboneFast on all switches running STP. It can be added without disruption to a production network. Issue this command in order to enable BackboneFast:

```
set spantree backbonefast enable
```

Note: This global level command needs to be configured on all switches in a domain as it adds functionality to the STP protocol that all switches need to understand.

Other Options

BackboneFast is not supported on 2900XLs and 3500s. It must not be enabled if the switch domain contains these switches in addition to Catalyst 4500/4000, 5500/5000, and 6500/6000 switches.

You do not need to configure BackboneFast with RSTP or IEEE 802.1w because the mechanism is natively included and automatically enabled in RSTP.

Spanning Tree Loop Guard

Loop guard is a Cisco proprietary optimization for STP. Loop guard protects L2 networks from loops that are caused by:

- Network interfaces that malfunction
- Busy CPUs
- Anything that prevents the normal forwarding of BPDUs

An STP loop occurs when a blocking port in a redundant topology erroneously transitions to the forwarding state. This transition usually happens because one of the ports in a physically redundant topology (not necessarily the blocking port) ceases to receive BPDUs.

Loop guard is only useful in switched networks where switches are connected by point-to-point links. Most modern campus and data center networks are these types of networks. On a point-to-point link, a designated bridge cannot disappear unless it sends an inferior BPDU or brings the link down. The STP loop guard feature was introduced in CatOS version 6.2(1) for Catalyst 4000 and Catalyst 5000 platforms, and in version 6.2(2) for the Catalyst 6000 platform.

Refer to Spanning-Tree Protocol Enhancements using Loop Guard and BPDU Skew Detection

Features for more information on loop guard.

Operational Overview

Loop guard checks to determine if a root port or an alternate/backup root port receives BPDUs. If the port does not receive BPDUs, loop guard puts the port into an inconsistent state (blocking) until the port starts to receive BPDUs again. A port in the inconsistent state does not transmit BPDUs. If such a port receives BPDUs again, the port (and link) is deemed viable again. The loop-inconsistent condition is removed from the port, and the STP determines the port state because such recovery is automatic.

Loop guard isolates the failure and lets spanning tree converge to a stable topology without the failed link or bridge. Loop guard prevents STP loops with the speed of the STP version in use. There is no dependency on STP itself (802.1d or 802.1w) or when the STP timers are tuned. For these reasons, implement loop guard in conjunction with UDLD in topologies that rely on STP and where the software supports the features.

When the loop guard blocks an inconsistent port, this message is logged:

```
set spantree backbonefast enable
```

When the BPDU is received on a port in a loop-inconsistent STP state, the port transitions into another STP state. In accordance with the received BPDU, the recovery is automatic, and no intervention is necessary. After the recovery, this message is logged.

```
set spantree backbonefast enable
```

Interaction with Other STP Features

- **Root guard**Root guard forces a port to be designated always. Loop guard is effective only if the port is the root port or an alternate port. These functions are mutually exclusive. Loop guard and root guard cannot be enabled on a port at the same time.
- **UplinkFast**Loop guard is compatible with UplinkFast. If loop guard puts a root port into a blocking state, UplinkFast puts a new root port into forwarding state. Also, UplinkFast does not select a loop-inconsistent port as a root port.
- **BackboneFast**Loop guard is compatible with BackboneFast. The reception of an inferior BPDU that comes from a designated bridge triggers BackboneFast. Because BPDUs are received from this link, loop guard is not activated, so BackboneFast and loop guard are compatible.
- **PortFast**PortFast transitions a port into the forwarding designated state immediately upon linkup. Because a PortFast-enabled port cannot be a root or alternate port, loop guard and PortFast are mutually exclusive.
- **PAgP**Loop guard uses the ports that are known to STP. Therefore, loop guard can take advantage of the abstraction of logical ports that PAgP provides. However, in order to form a channel, all the physical ports that are grouped in the channel must have compatible

configurations. PAgP enforces the uniform configuration of loop guard on all the physical ports to form a channel.**Note:** These are caveats when you configure loop guard on an EtherChannel:STP always picks the first operational port in the channel in order to send the BPDUs. If that link becomes unidirectional, loop guard blocks the channel, even if other links in the channel function properly.If ports which are already blocked by loop guard are grouped together in order to form a channel, STP loses all the state information for those ports. The new channel port can attain the forwarding state with a designated role.If a channel is blocked by loop guard and the channel breaks, STP loses all the state information. The individual physical ports can attain the forwarding state with designated role, even if one or more of the links that formed the channel are unidirectional.In the last two cases in this list, there is a possibility of a loop until UDLD detects the failure. But loop guard is not able to detect the loop.

Loop Guard and UDLD Feature Comparison

Loop guard functionality and UDLD functionality partially overlap. Both protect against STP failures that unidirectional links cause. But these two features are different in the approach to the problem and also in functionality. Specifically, there are certain unidirectional failures that UDLD cannot detect, such as failures that are caused by a CPU that does not send BPDUs. Additionally, the use of aggressive STP timers and RSTP mode can result in loops before UDLD can detect the failures.

Loop guard does not work on shared links or in situations in which the link has been unidirectional since the linkup. In the case that the link has been unidirectional since the linkup, the port never receives BPDUs and becomes designated. This behavior can be normal, so loop guard does not cover this particular case. UDLD does provide protection against such a scenario.

Enable both UDLD and loop guard in order to provide the highest level of protection. Refer to the *Loop Guard vs. Unidirectional Link Detection* section of Spanning-Tree Protocol Enhancements using Loop Guard and BPDU Skew Detection Features for a loop guard and UDLD feature comparison,.

Recommendation

Cisco recommends that you enable loop guard globally on a switch network with physical loops. In version 7.1(1) of the Catalyst software and later, you can enable loop guard globally on all ports. Effectively, the feature is enabled on all point-to-point links. The duplex status of the link detects the point-to-point link. If duplex is full, the link is considered point-to-point. Issue this command in order to enable global loop guard:

```
set spantree global-default loopguard enable
```

Other Options

For switches that do not support global loop guard configuration, enable the feature on all individual ports, which includes port channel ports. Although there are no benefits to enablement of loop guard on a designated port, this enablement is not an issue. In addition, a valid spanning

tree reconvergence can actually turn a designated port into a root port, which renders the feature useful on this port. Issue this command in order to enable loop guard:

```
set spantree guard loop mod/port
```

Networks with loop-free topologies can still benefit from loop guard in the case that loops are introduced accidentally. However, enablement of loop guard in this type of topology can lead to network isolation problems. In order to build loop-free topologies and avoid network isolation problems, issue these commands to disable loop guard globally or individually. Do not enable loop guard on shared links.

- ```
  set spantree global-default loopguard disable
  !--- This is the global default.
  ```
  or
- ```
  set spantree guard none mod/port
  !--- This is the default port configuration.
  ```

Spanning Tree Root Guard

The root guard feature provides a way to enforce the root bridge placement in the network. Root guard ensures that the port on which root guard is enabled is the designated port. Normally, root bridge ports are all designated ports, unless two or more ports of the root bridge are connected together. If the bridge receives superior STP BPDUs on a root guard-enabled port, the bridge moves this port to a root-inconsistent STP state. This root-inconsistent state is effectively equal to a listening state. No traffic is forwarded across this port. In this way, root guard enforces the position of the root bridge. Root guard is available in CatOS for Catalyst 29xx, 4500/4000, 5500/5000, and 6500/6000 in software version 6.1.1 and later.

Operational Overview

Root guard is an STP built-in mechanism. Root guard does not have a timer of its own, and it relies on the reception of BPDU only. When root guard is applied to a port, root guard does not allow a port to become a root port. If reception of a BPDU triggers a spanning tree convergence that makes a designated port become a root port, the port is put into a root-inconsistent state. This syslog message shows the action:

```
set spantree guard none mod/port
!--- This is the default port configuration.
```

After the port ceases to send superior BPDUs, the port is unblocked again. Through STP, the port goes from the listening state to the learning state, and eventually transitions to the forwarding state. Recovery is automatic, and no human intervention is necessary. This syslog message provides an example:

```
set spantree guard none mod/port
!--- This is the default port configuration.
```

Root guard forces a port to be designated and loop guard is effective only if the port is the root port or an alternate port. Therefore, the two functions are mutually exclusive. Loop guard and root guard cannot be enabled on a port at the same time.

Refer to Spanning Tree Protocol Root Guard Enhancement for more information.

Recommendation

Cisco recommends that you enable the root guard feature on ports that connect to network devices that are not under direct administrative control. In order to configure root guard, issue this command:

```
set spantree guard root mod/port
```

EtherChannel

EtherChannel technologies allow the inverse multiplexing of multiple channels (up to eight on Catalyst 6500/6000) into a single logical link. Although each platform differs from the next in implementation, it is important to understand the common requirements:

- An algorithm to statistically multiplex frames over multiple channels
- Creation of a logical port so that a single instance of STP can be run
- A channel management protocol such as PAgP or Link Aggregation Control Protocol (LACP)

Frame Multiplexing

EtherChannel encompasses a frame distribution algorithm that efficiently multiplexes frames across the component 10/100 or gigabit links. Differences in algorithms per platform arise from the capability of each type of hardware to extract frame header information in order to make the distribution decision.

The load distribution algorithm is a global option for both channel-control protocols. PAgP and LACP use the frame distribution algorithm because the IEEE standard does not mandate any particular distribution algorithms. However, any distribution algorithm ensures that, when frames are received, the algorithm does not cause the misordering of frames that are part of any given conversation or duplication of frames.

Note: This information must be considered:

- The Catalyst 6500/6000 has more recent switching hardware than the Catalyst 5500/5000 and can read IP Layer 4 (L4) information at wire rate in order to make a more intelligent multiplexing decision than simple MAC L2 information.
- The Catalyst 5500/5000 capabilities depend on the presence of an Ethernet Bundling Chip

(EBC) on the module. The **show port capabilities *mod/port*** command confirms what is possible for each port.

Refer to this table, which illustrates the frame distribution algorithm in detail for each listed platform:

Platform	Channel Load Balancing Algorithm
Catalyst 5500/5000 Series	A Catalyst 5500/5000 with the necessary modules allows two to four links to be present per FEC[1], though they must be on the same module. Source and destination MAC address pairs determine the link chosen for frame forwarding. An X-OR operation is performed on the least significant two bits of the source MAC address and the destination MAC address. This operation yields one of four results: (0 0), (0 1), (1 0), or (1 1). Each of these values points to a link in the FEC bundle. In the case of a two-port Fast EtherChannel, only a single bit is used in the X-OR operation. Circumstances can occur where one address in the source/destination pair is a constant. For example, the destination can be a server or, even more likely, a router. In that case, statistical load balancing is seen because the source address is always different.
Catalyst 4500/4000 Series	Catalyst 4500/4000 EtherChannel distributes frames across the links in a channel (on a single module) based on the low-order bits of the source and destination MAC addresses of each frame. In comparison with the Catalyst 5500/5000, the algorithm is more involved and uses a deterministic hash of these fields of the MAC DA (bytes 3, 5, 6), SA (bytes 3, 5, 6), ingress port, and VLAN ID. The frame distribution method is not configurable.
Catalyst 6500/6000 Series	There are two possible hashing algorithms, depending on the Supervisor Engine hardware. The hash is a seventeenth degree polynomial implemented in hardware that, in all cases, takes the MAC address, IP address, or IP TCP/UDP[2] port number and applies the algorithm to generate a three bit value. This is done separately for both source and destination addresses. The results are then XORd to generate another three-bit value that is used to determine which port in the channel is used to forward the packet. Channels on the Catalyst 6500/6000 can be formed between ports on any module and can be up to 8 ports.

[1] FEC = Fast EtherChannel

[2] UDP = User Datagram Protocol

This table indicates the distribution methods supported on the various Catalyst 6500/6000 Supervisor Engine models and their default behavior.

Hardware	Description	Distribution Methods
WS-F6020 (L2 Engine)	Early Supervisor Engine 1	L2 MAC: SA; DA; SA & DA
WS-F6020A (L2 Engine) WS-F6K-PFC (L3 Engine)	Later Supervisor Engine 1 and Supervisor Engine 1A/PFC1	L2 MAC: SA; DA; SA & DA L3 IP: SA; DA; SA and DA (default)
WS-F6K-PFC2	Supervisor Engine 2/PFC2 (needs CatOS 6.x)	L2 MAC: SA; DA; SA & DA L3 IP: SA; DA; SA & DA (default) L4 session: S port; D port; S & D port (default)
WS-F6K-PFC3BXL WS-F6K-PFC3B WS-F6K-PFC3A	Supervisor Engine 720/PFC3A (needs CatOS 8.1.x) Supervisor Engine 720/Supervisor Engine 32/PFC3B (needs CatOS 8.4.x) Supervisor Engine 720/PFC3BXL (needs CatOS 8.3.x)	L2 MAC: SA; DA; SA & DA L3 IP: SA; DA; SA & DA (default) L4 session: S port; D port; S & D port IP-VLAN-L4 session: SA & VLAN & S port; DA & VLAN & D port; SA & DA & VLAN & S port & D port

Note: With L4 distribution, the first fragmented packet uses L4 distribution. All subsequent packets use L3 distribution.

More details of EtherChannel support on other platforms and how to configure and troubleshoot them can be found in these documents:

- Understanding EtherChannel Load Balancing and Redundancy on Catalyst Switches
- Configure EtherChannel Between Catalyst 4500/4000, 5500/5000, and 6500/6000 Switches That Run CatOS System Software
- Configuring LACP (802.3ad) Between a Catalyst 6500/6000 and a Catalyst 4500/4000
- Configuring Layer 3 and Layer 2 EtherChannel

Recommendation

Catalyst 6500/6000 series switches perform load balancing by IP address by default. This is recommended in CatOS 5.5, assuming that IP is the dominant protocol. Issue this command in

order to set load balancing:

```
set port channel all distribution ip both
!--- This is the default.
```

Catalyst 4500/4000 and 5500/5000 series frame distribution by L2 MAC address is acceptable in most networks. However, the same link is used for all traffic if there are only two main devices that talk over a channel (as SMAC and DMAC are constant). This can typically be an issue for server back up and other large file transfers or for a transit segment between two routers.

Although the logical aggregate port (agport) can be managed by SNMP as a separate instance and aggregate throughput statistics gathered, Cisco still recommends that you manage each of the physical interfaces separately in order to check how the frame distribution mechanisms are working and whether statistical load balancing is being achieved.

A new command, the **show channel traffic** command, in CatOS 6.x can display percentage distribution statistics more easily than if you check individual port counters with the **show counters *mod/port*** command or the **show mac *mod/port*** command in CatOS 5.x. Another new command, the **show channel hash** command, in CatOS 6.x allows you to check, based on the distribution mode, which port would be selected as the outgoing port for certain addresses and/or port numbers. The equivalent commands for LACP channels are the **show lacp-channel traffic** command and the **show lacp-channel hash** command.

Other Options

These are possible steps to take if the relative limitations of Catalyst 4500/4000 or Catalyst 5500/5000 MAC-based algorithms are an issue, and good statistical load balancing is not achieved:

- Point-deploy Catalyst 6500/6000 switches
- Increase the bandwidth without channeling by switching, for example, from several FE ports to one GE port, or from several GE ports to one 10 GE port
- Re-address pairs of end stations with large volume flows
- Provision dedicated links/VLANs for high bandwidth devices

EtherChannel Configuration Guidelines and Restrictions

EtherChannel verifies port properties on all physical ports before it aggregates compatible ports into a single logical port. Configuration guidelines and restrictions vary for different switch platforms. Follow the guidelines in order to avoid bundling problems. For example, if QoS is enabled, EtherChannels do not form when bundling Catalyst 6500/6000 series switching modules with different QoS capabilities. In Cisco IOS Software, you can disable the QoS port attribute check on the EtherChannel bundling with the **no mls qos channel-consistency** port-channel interface command. An equivalent command in order to disable the QoS port attribute check is not available in CatOS. You can issue the **show port capability *mod/port*** command in order to display the QoS port capability and determine if ports are compatible.

Follow these guidelines for different platforms in order to avoid configuration problems:

- The *EtherChannel Configuration Guidelines* section of Configuring EtherChannel (Catalyst

6500/6000)
- The *EtherChannel Configuration Guidelines and Restrictions* section of Configuring Fast EtherChannel and Gigabit EtherChannel (Catalyst 4500/4000)
- The *EtherChannel Configuration Guidelines and Restrictions* section of Configuring Fast EtherChannel and Gigabit EtherChannel (Catalyst 5000)

Note: The maximum number of port channels that the Catalyst 4000 supports is 126. With software releases 6.2(1) and earlier, the six- and nine-slot Catalyst 6500 series switches support a maximum of 128 EtherChannels. In software release 6.2(2) and later releases, the spanning tree feature handles the port ID. Therefore, the maximum number of EtherChannels with support is 126 for a six- or nine-slot chassis and 63 for a 13-slot chassis.

Port Aggregation Protocol

PAgP is a management protocol that checks for parameter consistency at either end of the link and assists the channel in adapting to link failure or addition. Note these facts about PAgP:

- PAgP requires that all ports in the channel belong to the same VLAN or are configured as trunk ports. (Because dynamic VLANs can force the change of a port into a different VLAN, they are not included in EtherChannel participation.)
- When a bundle already exists and the configuration of one port is modified (such as changing VLAN or trunking mode), all ports in the bundle are modified to match that configuration.
- PAgP does not group ports that operate at different speeds or port duplex. If speed and duplex are changed when a bundle exists, PAgP changes the port speed and duplex for all ports in the bundle.

Operational Overview

The PAgP port controls each individual physical (or logical) port to be grouped. PAgP packets are sent using the same multicast group MAC address that is used for CDP packets, **01-00-0c-cc-cc-cc**. The protocol value is 0x0104. This is a summary of the protocol operation:

- As long as the physical port is `up`, PAgP packets are transmitted every second during detection and every 30 seconds in steady state.
- The protocol listens for PAgP packets that prove the physical port has a bidirectional connection to another PAgP-capable device.
- If data packets but no PAgP packets are received, it is assumed that the port is connected to a non-PAgP capable device.
- As soon as two PAgP packets have been received on a group of physical ports, it tries to form an aggregated port.
- If PAgP packets stop for a period, the PAgP state is torn `down`.

Normal Processing

These concepts must be defined to aid understanding of the protocol behavior:

- **Agport**—a logical port composed of all physical ports in the same aggregation, it can be identified by its own SNMP ifIndex. Therefore, an agport does not contain non-operational ports.

- **Channel**—an aggregation satisfying the formation criteria; it therefore could contain non-operational ports (agports are a subset of channels). Protocols including STP and VTP, but excluding CDP and DTP, run above PAgP over the agports. None of these protocols can send or receive packets until PAgP attaches their agports to one or more physical ports.
- **Group Capability**—each physical port and agport possesses a configuration parameter called the group-capability. A physical port can be aggregated with another physical port if and only if they have the same group-capability.
- **Aggregation Procedure**—when a physical port reaches the `UpData` or `UpPAgP` states, it is attached to an appropriate agport. When it leaves either of those states for another state, it is detached from the agport.

Definitions of the states and creation procedures are given in this table:

State	Meaning
`UpData`	No PAgP packets have been received. PAgP packets are sent. The physical port is the only one connected to its agport. Non-PAgP packets are passed in and out between physical port and agport.
`BiDir`	Exactly one PAgP packet has been received that proves a bidirectional connection exists to exactly one neighbor. The physical port is not connected to any agport. PAgP packets are sent and can be received.
`UpPAgP`	This physical port, perhaps in association with other physical ports, is connected to an agport. PAgP packets are sent and received on the physical port. Non-PAgP packets are passed in and out between physical port and agport.

Both ends of both connections must agree on what the grouping is going to be, defined as the largest group of ports in the agport that is permitted by both ends of the connection.

When a physical port reaches the `UpPAgP` state, it is assigned to the agport that has member physical ports that match the group-capability of the new physical port and that are in the `BiDir` or `UpPAgP` states. (Any such BiDir ports are moved to the `UpPAgP` state at the same time.) If there is no agport whose constituent physical port parameters are compatible with the newly ready physical port, it is assigned to an agport with suitable parameters that has no associated physical ports.

A PAgP timeout can occur on the last neighbor known on the physical port. The port timing out is removed from the agport. At the same time, all physical ports on the same agport whose timers have also timed out are removed. This enables an agport whose other end has died to be torn down all at once, instead of one physical port at a time.

Behavior in Failure

If a link in an existing channel is failed, (for example, port unplugged, Gigabit Interface Converter [GBIC] removed, or fiber broken), the agport is updated and the traffic is hashed over the remaining links within one second. Any traffic that does not need to be rehashed after the failure (traffic that continues to send on the same link) does not suffer any loss. Restoration of the failed

link triggers another update to the agport, and traffic is hashed again.

Note: The behavior when a link fails in a channel due to a power-off or the removal of a module can be different. By definition, there need to be two physical ports to a channel. If one port is lost from the system in a two-port channel, the logical agport is torn down and the original physical port is re-initialized with respect to Spanning Tree. This means traffic can be discarded until STP allows the port to become available to data again.

There is an exception to this rule on the Catalyst 6500/6000. In versions earlier than CatOS 6.3, an agport is not torn `down` during module removal if the channel is comprised of ports on modules 1 and 2 only.

This difference in the two failure modes is important when maintenance of a network is planned, as there can be an STP TCN to consider when performing an on-line removal or insertion of a module. As stated , it is important to manage each physical link in the channel with the NMS since the agport can remain undisturbed through a failure.

These are suggested steps in order to mitigate an unwanted topology change on the Catalyst 6500/6000:

- If a single port is used per module to form a channel, three or more modules must be used (three ports or more total).
- If the channel spans two modules, two ports on each module must be used (four ports total).
- If a two-port channel is needed across two cards, use only the Supervisor Engine ports.
- Upgrade to CatOS 6.3, which handles module removal without STP recalculation for channels split across modules.

Configuration Options

EtherChannels can be configured in different modes, as summarized in this table:

Mode	Configurable Options
`On`	PAgP not in operation. The port is channeling regardless of how the neighbor port is configured. If the neighbor port mode is on, a channel is formed.
`Off`	The port is not channeling regardless of how the neighbor is configured.
`Auto (default)`	Aggregation is under control of the PAgP protocol. Places a port into a passive `negotiating` state, and no PAgP packets are sent on the interface until at least one PAgP packet is received that indicates that the sender is operating in `desirable` mode.
`Desirable`	Aggregation is under control of the PAgP protocol. Places a port into an active `negotiating`

	state, in which the port initiates negotiations with other ports by sending PAgP packets. A channel is formed with another port group in either desirable or auto mode.
`Non-silent` (default on Catalyst 5500/5000 fiber FE and GE ports)	An `auto` or `desirable` mode keyword. If no data packets are received on the interface, then the interface is never attached to an agport and cannot be used for data. This bidirectionality check was provided for specific Catalyst 5500/5000 hardware as some link failures result in the channel being broken apart. Because `non-silent` mode is enabled, a recovering neighbor port is never allowed to come back up and break the channel apart unnecessarily. More flexible bundling and improved bidirectionality checks are present by default in Catalyst 4500/4000 and 6500/6000 series hardware.
`Silent` (default on all Catalyst 6500/6000 and 4500/4000 ports and 5500/5000 copper ports)	An `auto` or `desirable` mode keyword. If no data packets are received on the interface, after a 15 second timeout period, the interface is attached by itself to an agport and can thus be used for data transmission. `Silent` mode also allows for channel operation when the partner can be an analyzer or server that never sends PAgP.

The `silent/non-silent` settings affect how ports react to situations that cause unidirectional traffic or how they achieve fail-over. When a port is unable to transmit (because of a failed physical sublayer [PHY] or a broken fiber or cable, for example), this can still leave the neighbor port in an operational state. The partner continues to transmit data, but data is lost, as return traffic cannot be received. Spanning Tree loops can also form because of the unidirectional nature of the link.

Some fiber ports have the desired capability of bringing the port to a non-operational state when it loses its receive signal (FEFI). This causes the partner port to go non-operational and effectively causes the ports at both ends of the link to go down.

When using devices that transmit data (such as BPDUs) and cannot detect unidirectional conditions, `non-silent` mode must be used in order to allow the ports to remain non-operational until receive data is present and the link is verified to be bidirectional. The time it takes for PAgP to detect a unidirectional link is around 3.5 * 30 seconds = 105 seconds, where 30 seconds is the time between two successive PAgP messages. UDLD is recommended as a more rapid detector to uni-directional links.

When using devices that do not transmit any data, `silent` mode must be used. This forces the port to become connected and operational regardless of whether received data is present or not. Additionally, for those ports that can detect the presence of a unidirectional condition, such as newer platforms using L1 FEFI and UDLD, silent mode is used by default.

Verification

is table depicts a summary of all the possible PAgP channeling mode scenarios between two directly connected switches (Switch-A and Switch-B). Some of these combinations can cause STP to put the ports on the channeling side into the `errdisable` state (that is, some of the combinations shut down the ports on the channeling side).

Switch-A Channel Mode	Switch-B Channel Mode	Channel State
On	On	Channel (non-PAgP)
On	Off	Not Channel (errdisable)
On	Auto	Not Channel (errdisable)
On	Desirable	Not Channel (errdisable)
Off	On	Not Channel (errdisable)
Off	Off	Not Channel
Off	Auto	Not Channel
Off	Desirable	Not Channel
Auto	On	Not Channel (errdisable)
Auto	Off	Not Channel
Auto	Auto	Not Channel
Auto	Desirable	PAgP Channel
Desirable	On	Not Channel (errdisable)
Desirable	Off	Not Channel
Desirable	Auto	PAgP Channel
Desirable	Desirable	PAgP Channel

Recommendation

Cisco recommends that PAgP be enabled on all switch-to-switch channel connections, avoiding `on` mode. The preferred method is to set `desirable` mode at both ends of a link. The additional recommendation is to leave the `silent/non-silent` keyword at default - `silent` on Catalyst 6500/6000 and 4500/4000 switches, `non-silent` on Catalyst 5500/5000 fiber ports.

As discussed in this document, The explicit configuration of channeling off on all other ports is helpful for rapid data forwarding. Waiting up to 15 seconds for PAgP to timeout on a port that is not to be used for channeling must be avoided, especially since the port is then handed over to STP, which itself can take 30 seconds to allow data forwarding, plus potentially 5 seconds for DTP for a total of 50 seconds. The **set port host** command is discussed in more detail in the STP section of this document.

```
set port channel port range mode desirable

set port channel port range mode off
!--- Ports not channeled; part of the set port hostcommand.
```

This command assigns channels **an admin group** number, seen with a **show channel group** command. Addition and removal of channeling ports to the same agport can then be managed by the admin number if desired.

Other Options

Another common configuration for customers who have a model of minimal administration at the access layer is to set the mode to `desirable` at the distribution and core layers, and leave the access layer switches at the default `auto` configuration.

When channeling to devices that do not support PAgP, the channel needs to be hard-coded `on`. This applies to devices such as servers, Local Director, content switches, routers, switches with older software, Catalyst XL switches, and Catalyst 8540s. Issue this command:

```
set port channel port range mode on
```

The new 802.3ad IEEE LACP standard, available in CatOS 7.x, will likely supersede PAgP in the long term because it brings the benefit of cross-platform and vendor interoperability.

Link Aggregation Control Protocol

LACP is a protocol that allows ports with similar characteristics to form a channel through dynamic negotiation with adjoining switches. PAgP is a Cisco-proprietary protocol that can be run only on Cisco switches and those switches that are released by licensed vendors. But LACP, which is defined in IEEE 802.3ad, allows Cisco switches to manage Ethernet channeling with devices that conform to the 802.3ad specification. CatOS 7.x software releases introduced LACP support.

There is very little difference between LACP and PAgP from a functional perspective. Both protocols support a maximum of eight ports in each channel, and the same port properties are checked before the formation of the bundle. These port properties include:

- Speed
- Duplex
- Native VLAN
- Trunking type

The notable differences between LACP and PAgP are:

- LACP can run only on full-duplex ports, and LACP does not support half-duplex ports.
- LACP supports hot standby ports. LACP always tries to configure the maximum number of compatible ports in a channel, up to the maximum number that the hardware allows (eight

ports). If LACP is not able to aggregate all the ports that are compatible, all the ports that cannot be actively included in the channel are put in hot standby state and used only if one of the used ports fails. An example of a situation in which LACP cannot aggregate all the compatible ports is if the remote system has more-restrictive hardware limitations.

Note: In CatOS, the maximum number of ports that the same administrative key can be assigned is eight. In Cisco IOS Software, LACP tries to configure the maximum number of compatible ports in an EtherChannel, up to the maximum number that the hardware allows (eight ports). An additional eight ports can be configured as hot standby ports.

Operational Overview

The LACP controls each individual physical (or logical) port that is to be bundled. LACP packets are sent with use of the multicast group MAC address, **01-80-c2-00-00-02**. The type/field value is 0x8809 with a subtype of 0x01. Here is a summary of the protocol operation:

- The protocol relies on the devices to advertise their aggregation capabilities and state information. The transmissions are sent on a regular, periodic basis ***on each*** "aggregatable" link.
- As long as the physical port is up, LACP packets are transmitted every second during detection and every 30 seconds in steady state.
- The partners on an "aggregatable" link listen to the information that is sent within the protocol and decide what actions to take.
- Compatible ports are configured in a channel, up to the maximum number that the hardware allows (eight ports).
- The aggregations are maintained by the regular, timely exchange of up-to-date state information between the link partners. If the configuration changes (because of a link failure, for example), the protocol partners time out and take appropriate action on the basis of the new state of the system.
- In addition to periodic LACP data unit (LACPDU) transmissions, if there is a change to the state information, the protocol transmits an event-driven LACPDU to the partner. The protocol partners take the appropriate action on the basis of the new state of the system.

LACP Parameters

In order to allow LACP to determine if a set of links connect to the same system and if those links are compatible from the point of view of aggregation, the ability to establish these parameters is necessary:

- A globally unique identifier for each system that participates in link aggregationEach system that runs LACP must be assigned a priority that can be chosen either automatically or by the administrator. The default system priority is 32768. The system priority is mainly used in conjunction with the MAC address of the system in order to form the system identifier.
- A means of identification of the set of capabilities that are associated with each port and with each aggregator, as a given system understands themEach port in the system must be assigned a priority either automatically or by the administrator. The default is 128. The priority is used in conjunction with the port number in order to form the port identifier.
- A means of identification of a link aggregation group and its associated aggregatorThe ability of a port to aggregate with another is summarized by a simple 16-bit integer parameter that is

strictly greater than zero. This parameter is called the "key". Different factors determine each key, such as:The port physical characteristics, which include:Data rateDuplexityPoint-to-point or shared mediumConfiguration constraints that the network administrator establishesTwo keys are associated with each port:An administrative key—This key allows for the manipulation of key values by the management. A user can choose this key.An operational key—The system uses this key in order to form aggregations. A user cannot choose or directly change this key.The set of ports in a system that share the same operational key value are said to be members of the same key group.

If you have two systems and a set of ports with the same administrative key, each system tries to aggregate the ports. Each system starts from the port with the highest priority in the highest-priority system. This behavior is possible because each system knows its own priority, which either the user or the system has assigned, and its partner priority, which was discovered through LACP packets.

Behavior in Failure

Failure behavior for LACP is the same as the behavior for PAgP. If a link in an existing channel is failed, the agport is updated and the traffic is hashed over the remaining links within one second. A link can fail for these and other reasons:

- A port is unplugged
- A GBIC is removed
- A fiber is broken
- Hardware failure (interface or module)

Any traffic that does not need to be rehashed after the failure (traffic that continues to send on the same link) does not suffer any loss. Restoration of the failed link triggers another update to the agport, and traffic is hashed again.

Configuration Options

LACP EtherChannels can be configured in different modes, as this table summarizes:

Mode	Configurable Options
`On`	The link aggregation is forced to be formed without any LACP negotiation. The switch neither sends the LACP packet nor processes any incoming LACP packet. If the neighbor port mode is `on`, a channel is formed.
`Off`	The port is not channeling, regardless of how the neighbor is configured.
`Passive (default)`	This is similar to the `auto` mode in PAgP. The switch does not initiate the channel, but does understand incoming LACP packets. The peer (in `active` state) initiates negotiation by sending out an LACP packet. The switch receives and replies to the packet, and eventually forms the aggregation channel with the peer.

Active	This is similar to the desirable mode in PAgP. The switch initiates the negotiation in order to form an aglink. The link aggregate is formed if the other end runs in LACP active or passive mode.

Verification (LACP and LACP)

The table in this section depicts a summary of all the possible LACP channeling mode scenarios between two directly connected switches (Switch-A and Switch-B). Some of these combinations can cause STP to put the ports on the channeling side into the errdisable state. This means that some of the combinations shut down the ports on the channeling side.

Switch-A Channel Mode	Switch-B Channel Mode	Switch-A Channel State	Switch-B Channel State
On	On	Channel (non-LACP)	Channel (non-LACP)
On	Off	Not Channel (errdisable)	Not Channel
On	Passive	Not Channel (errdisable)	Not Channel
On	Active	Not Channel (errdisable)	Not Channel
Off	Off	Not Channel	Not Channel
Off	Passive	Not Channel	Not Channel
Off	Active	Not Channel	Not Channel
Passive	Passive	Not Channel	Not Channel
Passive	Active	LACP Channel	LACP Channel
Active	Active	LACP Channel	LACP Channel

Verification (LACP and PAgP)

The table in this section depicts a summary of all the possible LACP-to-PAgP channeling mode scenarios between two directly connected switches (Switch-A and Switch-B). Some of these combinations can cause STP to put the ports on the channeling side into the errdisable state. This means that some of the combinations shut down the ports on the channeling side.

Switch-A Channel Mode	Switch-B Channel Mode	Switch-A Channel State	Switch-B Channel State

On	On	Channel (non-LACP)	Channel (non-PAgP)
On	Off	Not Channel (errdisable)	Not Channel
On	Auto	Not Channel (errdisable)	Not Channel
On	Desirable	Not Channel (errdisable)	Not Channel
Off	On	Not Channel	Not Channel (errdisable)
Off	Off	Not Channel	Not Channel
Off	Auto	Not Channel	Not Channel
Off	Desirable	Not Channel	Not Channel
Passive	On	Not Channel	Not Channel (errdisable)
Passive	Off	Not Channel	Not Channel
Passive	Auto	Not Channel	Not Channel
Passive	Desirable	Not Channel	Not Channel
Active	On	Not Channel	Not Channel (errdisable)
Active	Off	Not Channel	Not Channel
Active	Auto	Not Channel	Not Channel
Active	Desirable	Not Channel	Not Channel

Recommendation

Cisco recommends that you enable PAgP on channel connections between Cisco switches. When you channel to devices that do not support PAgP but support LACP, enable LACP through the configuration of LACP `active` on both ends of the devices. If either end of the devices does not support LACP or PAgP, you need to hard code the channel to `on`.

- **set channelprotocol lacp** *module*

  On switches that run CatOS, all ports on a Catalyst 4500/4000 and a Catalyst 6500/6000 use channel protocol PAgP by default and, as such, do not run LACP. In order to configure ports to use LACP, you need to set the channel protocol on the modules to LACP. LACP and PAgP cannot run on the same module on switches that run CatOS.

- **set port lacp-channel** *port_range admin-key*

  An **admin key** (administrative key) parameter is exchanged in the LACP packet. A channel only forms between ports that have the same admin key. The **set port lacp-channel *port_range admin-key*** command assigns channels an admin key number. The **show lacp-channel group** command shows the number. The **set port lacp-channel *port_range admin-key*** command assigns the same admin key to all ports in the port range. The admin key is randomly assigned if a specific key is not configured. Then, you can refer to the admin key, if desired, in order to manage the addition and removal of channeling ports to the same agport.

- **`set port lacp-channel`** *`port_range`* **`mode active`**

  The **set port lacp-channel *port_range* mode active** command changes the channel mode to `active` for a set of ports that were previously assigned the same admin key.

Additionally, LACP utilizes a 30-second interval timer (Slow_Periodic_Time) after the LACP EtherChannels are established. The number of seconds before invalidation of received LACPDU information with the use of long timeouts (3 x Slow_Periodic_Time) is 90. Use UDLD, which is a more rapid detector of unidirectional links. You cannot adjust the LACP timers, and today you cannot configure the switches to use the fast PDU transmission (every second) in order to maintain the channel after the channel is formed.

Other Options

If you have a model of minimal administration at the access layer, a common configuration is to set the mode to `active` at the distribution and core layers. Leave the access layer switches at the default `passive` configuration.

Unidirectional Link Detection

UDLD is a Cisco proprietary, lightweight protocol that was developed to detect instances of unidirectional communications between devices. Although there are other methods to detect the bidirectional state of transmission media, like FEFI, there are certain instances in which the L1 detection mechanisms are not sufficient. These scenarios can result in any of these occurrences:

- The unpredictable operation of STP
- Incorrect or excessive flooding of packets
- The black holing of traffic

The UDLD feature is intended to address these fault conditions on fiber and copper Ethernet interfaces:

- Monitor physical cabling configurations and shutdown any miswired ports as `errdisable`.
- Protect against uni-directional links. When a uni-directional link is detected, due to media or port/interface malfunction, the affected port is shut down as `errdisable`, and a corresponding syslog message generated.
- Furthermore, UDLD aggressive mode checks that a link that was previously deemed bidirectional does not lose connectivity during congestion and become unusable. UDLD performs ongoing connectivity tests across the link. The primary purpose of UDLD aggressive mode is to avoid the black holing of traffic in certain failed conditions.

Spanning Tree, with its steady state unidirectional BPDU flow, was an acute sufferer from these failures. It is easy to see how a port can suddenly be unable to transmit BPDUs, causing an STP state change from `blocking` to `forwarding` on the neighbor. This change creates a loop, since the port is still able to receive.

Operational Overview

UDLD is a L2 protocol working above the LLC layer (destination MAC 01-00-0c-cc-cc-cc, SNAP HDLC protocol type 0x0111). When running UDLD in combination with FEFI and autonegotiation

L1 mechanisms, it is possible to validate the physical (L1) and logical (L2) integrity of a link.

UDLD has provisions for features and protection that FEFI and autonegotiation cannot perform, namely the detection and caching of neighbor information, the ability to shutdown any misconnected ports, and detect logical interface/port malfunctions or faults on links that are not point-to-point (those traversing media-converters or hubs).

UDLD employs two basic mechanisms; it learns about the neighbors, and keeps the information up-to-date in a local cache, and sends a train of UDLD probe/echo (hello) messages whenever it detects a new neighbor or whenever a neighbor requests a re-synchronization of the cache.

UDLD constantly sends probe messages on all ports on which UDLD is enabled. Whenever a specific "triggering" UDLD message is received on a port, a detection phase and validation process begins. If at the end of this process all valid conditions are met, the port state is not altered. In order to meet the conditions, the port must be bidirectional and correctly wired. Otherwise, the port is `errdisable`, and a syslog message displays. The syslog message is similar to these messages:

- `UDLD-3-DISABLE: Unidirectional link detected on port [dec]/[dec]. Port disabled`
- `UDLD-4-ONEWAYPATH: A unidirectional link from port [dec]/[dec] to port [dec]/[dec] of device [chars] was detected`

Refer to Messages and Recovery Procedures (Catalyst series switches, 7.6) for a complete list of system messages by facility, which includes UDLD events.

After a link is established and classed as bidirectional, UDLD continues to advertise probe/echo messages at a default interval of 15 seconds. This table represents valid UDLD link states as reported in the output of the **show udld port** command:

Port State	Comment
Undetermined	Detection in progress, or a neighboring UDLD entity has been disabled or its transmission has been blocked.
Not applicable	UDLD has been disabled.
Shutdown	Unidirectional link has been detected and the port disabled.
Bidirectional	Bidirectional link has been detected.

- **Neighbor Cache Maintenance**—UDLD periodically sends hello probe/echo packets on every active interface, in order to maintain the integrity of the UDLD neighbor cache. Whenever a hello message is received, it is cached and kept in memory for a maximum period defined as the hold-time. When the hold-time expires, the respective cache entry is aged out. If a new hello message is received within the hold-time period, the new one replaces the older entry and the corresponding time-to-live timer is reset.
- In order to maintain the integrity of the UDLD cache, whenever a UDLD-enabled interface gets disabled or a device is reset, all existing cache entries for the interfaces affected by the configuration change are cleared and UDLD transmits at least one message to inform respective neighbors to flush the corresponding cache entries.
- **Echo Detection Mechanism**—the echoing mechanism forms the basis of the detection algorithm. Whenever a UDLD device learns about a new neighbor or receives a

resynchronization request from an out-of-synch neighbor, it starts/restarts the detection window on its side of the connection and sends a burst of echo messages in reply. Since this behavior must be the same across all neighbors, the echo sender expects to receive echos back in reply. If the detection window ends and no valid reply message has been received, the link is considered unidirectional, and a link re-establishment or port shutdown process can be triggered.

Convergence Time

In order to prevent STP loops, CatOS 5.4(3) reduced the UDLD default message interval from 60 seconds to 15 seconds in order to shut down a unidirectional link before a blocked port was able to transition to a forwarding state.

Note: The message interval value determines the rate at which a neighbor sends UDLD probes after the linkup or detection phase. The message interval does not need to match on both ends of a link, although consistent configuration is desirable where possible. When UDLD neighbors are established, the configured message interval is sent and the timeout interval for that peer is calculated to be (3 * message_interval). Therefore, a peer relationship times out after three consecutive hellos (or probes) are missed. With the message intervals different on each side, this timeout value is different on each side.

The approximate time that is necessary for UDLD to detect a unidirectional failure is approximately (2.5 * message_interval + 4 seconds), or about 41 seconds with use of the default message interval of 15 seconds. This is well below the 50 seconds that are usually necessary for STP to reconverge. If the NMP CPU has some spare cycles and if you carefully monitor its utilization level, you can reduce the message interval (even) to the minimum of 7 seconds. This message interval helps speed up the detection by a significant factor.

Therefore, UDLD has an assumed dependency on default spanning tree timers. If you tune STP to converge more rapidly than UDLD, consider an alternate mechanism, such as the CatOS 6.2 loop guard feature. Also consider an alternate mechanism when you implement RSTP (IEEE 802.1w) because RSTP has convergence characteristics in the milliseconds, which depends on the topology. For these instances, use loop guard in conjunction with UDLD, which provides the most protection. Loop guard prevents STP loops with the speed of the STP version that is in use, and UDLD detects unidirectional connections on individual EtherChannel links or in cases in which BPDUs do not flow along the broken direction.

Note: UDLD does not catch every STP failure situation, such as failures that are caused by a CPU that does not send BPDUs for a time greater than (2 * FwdDelay + Maxage). For this reason, Cisco recommends that you implement UDLD in conjunction with loop guard (which was introduced in CatOS 6.2) in topologies that rely on STP.

Caution: Beware of earlier releases of UDLD that use a nonconfigurable 60-second default message interval. These releases are susceptible to spanning-tree loop conditions.

UDLD Aggressive Mode

Aggressive UDLD was created in order to specifically address those (few) cases in which an ongoing test of bidirectional connectivity is necessary. As such, the aggressive mode feature provides enhanced protection against dangerous unidirectional link conditions in these situations:

- When the loss of UDLD PDUs is symmetrical and both ends time out, neither port is errdisabled.
- One side of a link has a port stuck (both transmit [Tx] and Rx).
- One side of a link remains up while the other side of the link has gone down.
- Autonegotiation, or another L1 fault-detection mechanism, is disabled.
- A reduction of the reliance on L1 FEFI mechanisms is desirable.
- Maximum protection against unidirectional link failures on point-to-point FE/GE links is necessary. Specifically, where no failure between two neighbors is admissible, UDLD-aggressive probes can be considered as a “heartbeat”, the presence of which guarantees the health of the link.

The most common case for an implementation of aggressive UDLD is in order to perform the connectivity check on a member of a bundle when autonegotiation or another L1 fault-detection mechanism is disabled or unusable. This is particularly true with EtherChannel connections because PAgP/LACP, even if enabled, do not use very low hello timers at steady state. In this case, aggressive UDLD has the added benefit of prevention of possible spanning-tree loops.

The circumstances that contribute to the symmetrical loss of UDLD probe packets are more difficult to characterize. You must understand that normal UDLD does check for a unidirectional link condition, even after a link reaches bidirectional status. The intention of UDLD is to detect L2 problems that cause STP loops, and those problems are usually unidirectional because BPDUs flow only in one direction at steady state. Therefore, the use of normal UDLD in conjunction with autonegotiation and loop guard (for networks that rely on STP) is almost always sufficient. However, UDLD aggressive mode is beneficial in situations in which congestion is equally affected in both directions, which causes the loss of UDLD probes in both directions. For example, this loss of UDLD probes can occur if CPU utilization on each end of the link is elevated. Other examples of bidirectional loss of connectivity include the fault of one of these devices:

- A Dense Wavelength Division Multiplexing (DWDM) transponder
- A media converter
- A hub
- Another L1 device**Note:** The fault cannot be detected by autonegotiation.

Aggressive UDLD error disables the port in these failure situations. Consider the ramifications carefully when you enable UDLD aggressive mode on links that are not point-to-point. Links with media converters, hubs, or similar devices are not point-to-point. Intermediate devices can prevent the forwarding of UDLD packets and force a link to be shut down unnecessarily.

After all the neighbors of a port have aged out, UDLD aggressive mode (if it is enabled) restarts the linkup sequence in an effort to resynchronize with any potentially out-of-sync neighbors. This effort takes place in either the advertisement or the detection phase. If after a fast train of messages (eight failed retries) the link is still deemed "undetermined", the port is then put into `errdisable` state.

Note: Some switches are not aggressive UDLD-capable. Currently, the Catalyst 2900XL and Catalyst 3500XL have hard-coded message intervals of 60 seconds. This interval is not considered sufficiently fast to protect against potential STP loops (with use of the default STP parameters).

UDLD on Routed Links

For the purpose of this discussion, a routed link is one of two connection types:

- Point-to-point between two router nodesThis link is configured with a 30-bit subnet mask.
- A VLAN with multiple ports but that supports only routed connectionsAn example is a split L2 core topology.

Each Interior Gateway Routing Protocol (IGRP) has unique characteristics with respect to how it handles neighbor relationships and route convergence. The characteristics, which this section discusses, are relevant when you contrast two of the more prevalent routing protocols that are used today, Open Shortest Path First (OSPF) Protocol and Enhanced IGRP (EIGRP).

First, note that an L1 or L2 failure on any point-to-point routed network results in the almost immediate teardown of the L3 connection. Because the only switch port in that VLAN transitions to a not-connected state upon the L1/L2 failure, the MSFC auto-state feature synchronizes the L2 and L3 port states in approximately two seconds. This synchronization places the L3 VLAN interface in an up/down state (with the line protocol down).

Assume default timer values. OSPF sends hello messages every 10 seconds and has a dead interval of 40 seconds (4 * hello). These timers are consistent for OSPF point-to-point and broadcast networks. Because OSPF requires two-way communication in order to form an adjacency, the worst-case failover time is 40 seconds. This failover is the case even if the L1/L2 failure is not pure on a point-to-point connection, which leaves a half-operational scenario with which the L3 protocol must deal. Because the detection time of UDLD is very similar to the time of an OSPF dead timer that expires (about 40 seconds), the advantages of configuration of UDLD normal mode on an OSPF L3 point-to-point link are limited.

In many cases, EIGRP converges faster than OSPF. However, you must note that two-way communication is not necessary in order for neighbors to exchange routing information. In very specific half-operational failure scenarios, EIGRP is vulnerable to the black holing of traffic that lasts until some other event makes the routes by way of that neighbor "active". UDLD normal mode can alleviate the circumstances that this section notes. UDLD normal mode detects the unidirectional link failure and error disables the port.

For L3-routed connections that use any routing protocol, UDLD normal still provides protection against issues upon initial link activation. Such issues include miscabling or faulty hardware. Additionally, UDLD aggressive mode provides these advantages on L3-routed connections:

- Prevents the unnecessary black holing of traffic**Note:** Minimum timers are required in some cases.
- Places a flapping link into the `errdisable` state
- Protects against loops that result from L3 EtherChannel configurations

Default Behavior of UDLD

UDLD is disabled globally and enabled in readiness on fiber ports by default. Because UDLD is an infrastructure protocol that is necessary between switches only, UDLD is disabled by default on copper ports. Copper ports tend to be used for host access.

Note: UDLD must be enabled globally and at the interface level before neighbors can achieve bidirectional status. In CatOS 5.4(3) and later, the default message interval is 15 seconds and is configurable between 7 and 90 seconds.

Errdisable recovery is globally disabled by default. After it is enabled globally, if a port goes into `errdisable` state, the port is re-enabled automatically after a selected time interval. The default time

is 300 seconds, which is a global timer and maintained for all ports in a switch. You can manually prevent a port re-enablement if you set the errdisable timeout for that port to `disable`. Issue the **set port errdisable-timeout *mod/port* disable** command.

Note: Use of this command depends on your software version.

Consider use of the errdisable timeout feature when you implement UDLD aggressive mode with no out-of-band network management capabilities, particularly in the access layer or on any device that can become isolated from the network in the event of an errdisable situation.

Refer to Configuring Ethernet, Fast Ethernet, Gigabit Ethernet, and 10-Gigabit Ethernet Switching for more details on how to configure a timeout period for ports that are in the `errdisable` state.

Recommendation

Normal mode UDLD is sufficient in the vast majority of cases if you use it properly and in conjunction with the appropriate features and protocols. These features/protocols include:

- FEFI
- Autonegotiation
- Loop guard

When you deploy UDLD, consider if an ongoing test of bidirectional connectivity (aggressive mode) is necessary. Typically, if autonegotiation is enabled, aggressive mode is not necessary because autonegotiation compensates for the fault detection at L1.

Cisco recommends the enablement of UDLD normal mode on all point-to-point FE/GE links between Cisco switches in which the UDLD message interval is set to the 15-second default. This configuration assumes the default 802.1d spanning tree timers. Additionally, use UDLD in conjunction with loop guard in networks that rely on STP for redundancy and convergence. This recommendation applies to networks in which there are one or more ports in the STP blocking state in the topology.

Issue these commands in order to enable UDLD:

```
set udld enable
!--- After global enablement, all FE and GE fiber !--- ports have UDLD enabled by
default. set udld enable port range
!--- This is for additional specific ports and copper media, if needed.
```

You must manually enable ports that are error disabled because of unidirectional link symptoms. Issue the **set port enable** command.

Refer to Understanding and Configuring the Unidirectional Link Detection Protocol (UDLD) Feature for more details.

Other Options

For maximum protection against symptoms that result from unidirectional links, configure aggressive mode UDLD:

```
set udld aggressive-mode enable port_range
```

Additionally, you can tune the UDLD message interval value between 7 and 90 seconds at each end, where supported, for faster convergence:

```
set udld interval time
```

Consider use of the errdisable timeout feature on any device that can become isolated from the network in the event of an errdisable situation. This situation is typically true of the access layer and when you implement UDLD aggressive mode with no out-of-band network management capabilities.

If a port is placed in `errdisable` state, the port remains down by default. You can issue this command, which re-enables ports after a timeout interval:

Note: The timeout interval is 300 seconds by default.

```
>set errdisable-timeout enable ?
bpdu-guard
!--- This is BPDU port-guard. channel-misconfig !--- This is a channel
misconfiguration. duplex-mismatch udld other !--- These are other reasons. all !---
Apply errdisable timeout to all reasons.
```

If the partner device is not UDLD-capable, such as an end host or router, do not run the protocol. Issue this command:

```
set udld disable port_range
```

Test and Monitor UDLD

UDLD is not easy to test without a genuinely faulty/unidirectional component in the lab, such as a defective GBIC. The protocol was designed to detect less-common failure scenarios than those scenarios that are usually employed in a lab. For example, if you perform a simple test and unplug one strand of a fiber in order to see the desired `errdisable` state, you need to have turned off L1 autonegotiation. Otherwise, the physical port goes down, which resets UDLD message communication. The remote end moves to the undetermined state in UDLD normal. If you use UDLD aggressive mode, the remote end moves to the `errdisable` state.

There is an additional test method to simulate neighbor PDU loss for UDLD. Use MAC-layer filters in order to block the UDLD/CDP hardware address but allow other addresses to pass.

In order to monitor UDLD, issue these commands:

```
>show udld

UDLD              : enabled
Message Interval  : 15 seconds


>show udld port 3/1

UDLD              : enabled
Message Interval  : 15 seconds
Port      Admin Status  Aggressive Mode  Link State
--------  ------------  ---------------  ----------------
3/1       enabled       disabled         bidirectional
```

Also from `enable` mode, you can issue the hidden **show udld neighbor** command in order to check the UDLD cache contents (in the way that CDP does). A comparison of the UDLD cache to the CDP cache in order to verify if there is a protocol-specific anomaly is often useful. Whenever CDP is also affected, all PDUs/BPDUs are typically affected. Therefore, check STP also. For example, check for recent root identity changes or root/designated port placement changes.

```
>show udld neighbor 3/1

Port  Device Name            Device ID    Port-ID OperState
----- ---------------------  ------------ ------- ----------
3/1   TSC07117119M(Switch)   000c86a50433 3/1     bidirectional
```

Furthermore, you can monitor the UDLD status and configuration consistency with use of the Cisco UDLD SNMP MIB variables.

Jumbo Frame

The default Maximum Transmission Unit (MTU) frame size is 1518 bytes for all Ethernet ports, which includes GE and 10 GE. The jumbo frame feature enables interfaces to switch frames that are larger than the standard Ethernet frame size. The feature is useful in order to optimize server-to-server performance and to support applications such as Multi-Protocol Label Switching (MPLS), 802.1Q tunneling, and L2 Tunneling Protocol Version 3 (L2TPv3), which increase the size of the original frames.

Operational Overview

The IEEE 802.3 standard specification defines a maximum Ethernet frame size of 1518 bytes for regular frames and 1522 bytes for 802.1Q encapsulated frames. The 802.1Q encapsulated frames are sometimes referred to as "baby giants". In general, packets are classified as giant frames when the packets exceed the specified Ethernet maximum length for a specific Ethernet connection. Giant packets are also known as jumbo frames.

There are various reasons why the MTU size of certain frames can exceed 1518 bytes. These are some of the examples:

- Vendor-specific requirements—Applications and certain NICs can specify an MTU size that is outside of the standard 1500 bytes. The tendency to specify such MTU sizes is because of studies that have been undertaken, which prove that an increase in the size of an Ethernet frame can increase the average throughput.

- Trunking—In order to carry VLAN ID information between switches or other network devices, trunking has been employed to augment the standard Ethernet frame. Today, the two most common forms of trunking are the Cisco proprietary ISL encapsulation and IEEE 802.1Q.
- MPLS—After MPLS is enabled on an interface, it has the potential to augment the frame size of a packet. This augmentation depends on the number of labels in the label stack for an MPLS-tagged packet. The total size of a label is 4 bytes. The total size of a label stack is *n* x 4 bytes. If a label stack is formed, the frames can exceed the MTU.
- 802.1Q tunneling—802.1Q tunneling packets contain two 802.1Q tags, of which only one tag at a time is usually visible to the hardware. Therefore, the internal tag adds 4 bytes to the MTU value (payload size).
- Universal Transport Interface (UTI)/L2TPv3—UTI/L2TPv3 encapsulates L2 data that are to be forwarded over the IP network. The encapsulation can increase the original frame size by up to 50 bytes. The new frame includes a new IP header (20-byte), an L2TPv3 header (12-byte), and a new L2 header. The L2TPv3 payload consists of the complete L2 frame, which includes the L2 header.

The ability of the different Catalyst switches to support various frame sizes depends on many factors, which include the hardware and software. Certain modules can support larger frame sizes than others, even within the same platform.

- The Catalyst 5500/5000 switches provide support for jumbo frame in the CatOS 6.1 release. When the jumbo frames feature is enabled on a port, the MTU size increases to 9216 bytes. On 10/100-Mbps unshielded twisted pair (UTP)-based line cards, the maximum frame size that is supported is only 8092 bytes. This limitation is an ASIC limitation. There are generally no restrictions in the enablement of the jumbo frame size feature. You can use this feature with trunking/nontrunking and channeling/nonchanneling.
- The Catalyst 4000 switches (Supervisor Engine 1 [WS-X4012] and Supervisor Engine 2 [WS-X4013]) do not support jumbo frames because of an ASIC limitation. However, the exception is 802.1Q trunking.
- The Catalyst 6500 series platform can support jumbo frame sizes in CatOS release 6.1(1) and later. However, this support is dependent on the type of line cards that you use. There are generally no restrictions in the enablement of the jumbo frame size feature. You can use this feature with trunking/nontrunking and channeling/nonchanneling. The default MTU size is 9216 bytes after jumbo frame support has been enabled on the individual port. The default MTU is not configurable with use of CatOS. However, Cisco IOS Software Release 12.1(13)E introduced the **system jumbomtu** command in order to override the default MTU.

Refer to Jumbo/Giant Frame Support on Catalyst Switches Configuration Example for more information.

This table describes the MTU sizes that are supported by different line cards for Catalyst 6500/6000 series switches:

Note: The MTU size or packet size refers only to Ethernet payload.

Line Card	**MTU Size**
Default	9216 bytes
WS-X6248-RJ-45, WS-X6248A-RJ-45 WS-X6248-TEL, WS-X6248A-TEL WS-X6348-RJ-	8092 bytes

45(V), WS-X6348-RJ-21(V)	(limited by the PHY chip)
WS-X6148-RJ-45(V), WS-X6148-RJ-21(V) WS-X6148-45AF, WS-X6148-21AF	9100 bytes (@ 100 Mbps) 9216 bytes (@ 10 Mbps)
WS-X6148A-RJ-45, WS-X6148A-45AF, WS-X6148-FE-SFP	9216 bytes
WS-X6324-100FX-MM, -SM, WS-X6024-10FL-MT	9216 bytes
WS-X6548-RJ-45, WS-X6548-RJ-21, WS-X6524-100FX-MM WS-X6148X2-RJ-45, WS-X6148X2-45AF WS-X6196-RJ-21, WS-X6196-21AF WS-X6408-GBIC, WS-X6316-GE-TX , WS-X6416-GBIC WS-X6516-GBIC, WS-X6516A-GBIC, WS-X6816-GBIC Uplinks of Supervisor Engine 1, 2, 32 and 720	9216 bytes
WS-X6516-GE-TX	8092 bytes (@ 100 Mbps) 9216 bytes (@ 10 or 1000 Mbps)
WS-X6148-GE-TX, WS-X6148V-GE-TX, WS-X6148-GE-45AF, WS-X6548-GE-TX, WS-X6548V-GE-TX, WS-X6548-GE-45AF	1500 bytes (jumbo frame not supported)
WS-X6148A-GE-TX, WS-X6148A-GE-45AF, WS-X6502-10GE, WS-X67xx Series	9216 bytes
OSM ATM (OC12c)	9180 bytes
OSM CHOC3, CHOC12, CHOC48, CT3	9216 bytes (OCx and DS3) 7673 bytes (T1/E1)
Flex WAN	7673 bytes (CT3

	T1/DS0) 9216 bytes (OC3c POS) 7673 bytes (T1)
CSM (WS-X6066-SLB-APC)	9216 bytes (as of CSM 3.1(5) and 3.2(1))
OSM POS OC3c, OC12c, OC48c; OSM DPT OC48c, OSM GE WAN	9216 bytes

Layer 3 Jumbo Frame Support

With CatOS that runs on the Supervisor Engine and Cisco IOS Software that runs on the MSFC, the Catalyst 6500/6000 switches also provide L3 jumbo frame support in Cisco IOS® Software Release 12.1(2)E and later with the use of PFC/MSFC2, PFC2/MSFC2, or later hardware. If both ingress and egress VLANs are configured for jumbo frames, all the packets are hardware switched by the PFC at wire speed. If the ingress VLAN is configured for jumbo frames and the egress VLAN is not configured, there are two scenarios:

- A jumbo frame that is sent by the end host with the Don't Fragment (DF) bit set (for path MTU discovery)—The packet is dropped and an Internet Control Message Protocol (ICMP) unreachable is sent to the end host with the message code `fragment needed and DF set`.
- A jumbo frame that is sent by the end host with the DF bit not set—Packets are punted to MSFC2/MSFC3 to be fragmented and switched in software.

This table summarizes the L3 jumbo support for various platforms:

L3 Switch or Module	**Maximum L3 MTU Size**
Catalyst 2948G-L3/4908G-L3 Series	Jumbo frames are not supported.
Catalyst 5000 RSM[1]/RSFC[2]	Jumbo frames are not supported.
Catalyst 6500 MSFC1	Jumbo frames are not supported.
Catalyst 6500 MSFC2 and later	Cisco IOS Software Release 12.1(2)E: 9216 bytes

[1] RSM = Route Switch Module

[2] RSFC = Route Switch Feature Card

Network Performance Consideration

The performance of TCP over WANs (the Internet) has been extensively studied. This equation explains how TCP throughput has an upper bound that is based on:

- The Maximum Segment Size (MSS), which is the MTU length minus the length of the TCP/IP headers
- The Round Trip Time (RTT)
- The packet loss

According to this formula, the maximum TCP throughput that is achievable is directly proportional to the MSS. With constant RTT and packet loss, you can double the TCP throughput if you double the packet size. Similarly, when you use jumbo frames instead of 1518-byte frames, a six-fold increase in size can yield a potential six-fold improvement in the TCP throughput of an Ethernet connection.

Secondly, the ever-increasing performance demands of server farms require a more efficient means to ensure higher data rates with Network File System (NFS) UDP datagrams. NFS is the most widely deployed data storage mechanism to transfer files between UNIX-based servers, and it features 8400-byte datagrams. Given the extended 9 KB MTU of Ethernet, a single jumbo frame is large enough to carry an 8 KB application datagram (for example, NFS) plus the packet header overhead. This capability incidentally allows for more efficient direct memory access (DMA) transfers on the hosts because software does not need any more in order to fragment NFS blocks into separate UDP datagrams.

Recommendation

When you want jumbo frame support, constrain the use of jumbo frames to areas of the network where all switch modules (L2) and interfaces (L3) support jumbo frames. This configuration prevents fragmentation anywhere in the path. The configuration of jumbo frames that are larger than the supported frame length in the path eliminates any gains that are achieved by the use of the feature because fragmentation is required. As the tables in this Jumbo Frame section show, different platforms and line cards can vary with regard to the maximum packet sizes that are supported.

Configure jumbo frame-aware host devices with an MTU size that is the minimum common denominator that is supported by network hardware, for the entire L2 VLAN where the host device resides. In order to enable the jumbo frame support for modules with jumbo frame support, issue this command:

```
set port jumbo mod/port enable
```

In addition, if you desire jumbo frame support across L3 boundaries, configure the largest available MTU value of 9216 bytes on all the applicable VLAN interfaces. Issue the **mtu** command under the VLAN interfaces:

```
interface vlan vlan# mtu 9216
```

This configuration ensures that the L2 jumbo frame MTU that is supported by the modules is

always less than, or equal to, the value that is configured for the L3 interfaces that the traffic traverses. This prevents fragmentation when traffic is routed from the VLAN across the L3 interface.

Management Configuration

Considerations to help control, provision, and troubleshoot a Catalyst network are discussed in this section.

Network Diagrams

Clear network diagrams are a fundamental part of network operations. They become critical during troubleshooting and are the single most important vehicle for the communication of information when escalated to vendors and partners during an outage. Their preparation, readiness, and accessibility must not be underestimated.

Recommendation

Cisco recommends that you create these three diagrams:

- **Overall Diagram**—even for the largest networks, a diagram that shows the end-to-end physical and logical connectivity is important. It can be common for enterprises that have implemented a hierarchical design to document each layer separately. During planning and problem solving, however, it is often a good knowledge of how the domains link together that matters.
- **Physical Diagram**—shows all switch and router hardware and cabling. Trunks, links, speeds, channel groups, port numbers, slots, chassis types, software, VTP domains, root bridge, backup root bridge priority, MAC address, and blocked ports per VLAN must be labeled. It is often clearer to depict internal devices, such as the Catalyst 6500/6000 MSFC, as a router on a stick connected by way of a trunk.
- **Logical Diagram**—shows only L3 functionality (routers as objects, VLANs as Ethernet segments). IP addresses, subnets, secondary addressing, HSRP active and standby, access-core-distribution layers, and routing information must be labeled.

In-Band Management

Depending on the configuration, the switch in-band (internal) management interface (known as sc0) could have to handle this data:

- Switch management protocols such as SNMP, Telnet, Secure Shell Protocol (SSH), and syslog
- User data such as broadcasts and multicasts
- Switch control protocols such as STP BPDUs, VTP, DTP, CDP, and so on

It is common practice in the Cisco multilayer design to configure a management VLAN that spans a switched domain and contains all sc0 interfaces. This helps separate management traffic from user traffic and increases security of the switch management interfaces. This section describes the significance and potential problems of using the default VLAN 1 and running management traffic to the switch in the same VLAN as user traffic.

Operational Overview

The primary concern over the use of VLAN 1 for user data is that the Supervisor Engine NMP in general does not need to be interrupted by much of the multicast and broadcast traffic that is generated by end-stations. Older Catalyst 5500/5000 hardware, the Supervisor Engine I and Supervisor Engine II in particular, has limited resources for dealing with this traffic, though the principle applies to all Supervisor Engines. If the Supervisor Engine CPU, buffer, or in-band channel to the backplane is fully occupied listening to unnecessary traffic, it is possible that control frames can be missed. In a worst-case scenario, this could lead to a Spanning Tree loop or EtherChannel failure.

If the **show interface** and **show ip stats** commands are issued on the Catalyst, they can give some indication of the proportion of broadcast to unicast traffic and the proportion of IP to non-IP traffic (not typically seen in management VLANs).

A further health check for older Catalyst 5500/5000 hardware is to examine the output of **show *inband* | *biga*** (hidden command) for resource errors (RscrcErrors), similar to buffer drops in a router. If these resource errors go up continuously, memory is not available to receive system packets, perhaps because of a significant amount of broadcast traffic in the management VLAN. A single resource error can mean that the Supervisor Engine is unable to process a packet such as BPDUs, which could quickly become a problem because protocols such as spanning tree do not re-send missed BPDUs.

Recommendation

As highlighted in the Cat Control section of this document, VLAN 1 is a special VLAN that tags and handles most of the control plane traffic. VLAN 1 is enabled on all trunks by default. With larger campus networks, care needs to be taken about the diameter of the VLAN 1 **STP domain**; instability in one part of the network could affect VLAN 1, thereby influencing control-plane stability and therefore STP stability for all other VLANs. In CatOS 5.4 and later, it has been possible to limit VLAN 1 from carrying user data and running STP with this command:

```
clear trunk mod/port vlan 1
```

This does not stop control packets being sent from switch to switch in VLAN 1, as seen with a network analyzer. However, no data is forwarded, and STP is not be run over this link. Therefore, this technique can be used to break VLAN 1 up into smaller failure domains.

Note: It is not currently possible to clear VLAN 1 trunks on 3500s and 2900XLs.

Even if care has been taken with the campus design to constrain user VLANs to relatively small switch domains and correspondingly small failure/L3 boundaries, some customers are still tempted to treat the management VLAN differently and try to cover the whole network with a single management subnet. There is no technical reason that a central NMS application must be L2-adjacent to the devices it manages, nor is this a qualified security argument. Cisco recommends that you limit the diameter of the management VLANs to the same routed domain structure as user VLANs and considering out-of-band management and/or CatOS 6.x SSH support as a way to increase network management security.

Other Options

However, there are design considerations for these Cisco recommendations in some topologies. For example, a desirable and common Cisco multilayer design is one that avoids the use of an active Spanning Tree. This requires that you constrain each IP subnet/VLAN to a single access-layer switch, or cluster of switches. In these designs, there could be no trunking configured down to the access layer.

There is no easy answer to the question of whether a separate management VLAN be created and trunking enabled in order to carry it between L2 access and L3 distribution layers. These are two options for design review with your Cisco engineer:

- **Option 1:** trunk two or three unique VLANs from the distribution layer down to each access-layer switch. This allows for a data VLAN, a voice VLAN, and a management VLAN, for example, and still has the benefit that STP is inactive. (Note that if VLAN 1 is cleared from the trunks, there is an extra configuration step.) In this solution, there are also design points to consider in order to avoid the temporary black-holing of routed traffic during failure recovery: STP PortFast for trunks (CatOS 7.x and later) or VLAN Autostate synchronization with STP forwarding (later than CatOS 5.5[9]).
- **Option 2:** a single VLAN for data and management could be acceptable. With newer switch hardware, such as more powerful CPUs and control-plane rate-limiting controls, plus a design with relatively small broadcast domains as advocated by the multilayer design, the reality for many customers is that keeping the sc0 interface separate from the user data is less of an issue than it once was. A final decision is probably best taken with the examination of the broadcast traffic profile for that VLAN and a discussion of the capabilities of the switch hardware with your Cisco engineer. If the management VLAN does indeed contain all users on that access-layer switch, the use of IP input filters is highly recommended to secure the switch from users, as discussed in the Security Configuration section of this document.

Out-of-Band Management

Taking the arguments of the previous section one step further, network management can be made more highly available with the construction of a separate management infrastructure around the production network so that devices are always reachable remotely no matter what traffic-driven or control-plane events occur. These two approaches are typical:

- Out-of-Band Management with an exclusive LAN
- Out-of-Band Management with terminal servers

Operational Overview

Every router and switch in the network can be provided with an out-of-band Ethernet management interface on a management VLAN. One Ethernet port on each device is configured in the management VLAN and cabled outside the production network to a separate switched management network through the sc0 interface. Note that Catalyst 4500/4000 switches have a special me1 interface on the Supervisor Engine that is to be used for out-of-band management only, not as a switch port.

In addition, terminal server connectivity can be achieved through the configuration a Cisco 2600 or 3600 with RJ-45-to-serial cables to access the console port of every router and switch in the

layout. A terminal server also avoids the need for the configuration of backup scenarios, such as modems on auxiliary ports for every device. A single modem can be configured on the auxiliary port of the terminal server to provide dial-up service to the other devices during a network connectivity failure.

Recommendation

With this arrangement, two out-of-band paths to every switch and router are possible in addition to numerous in-band paths, thus enabling highly-available network management. Out-of-band is responsible for:

- Out-of-band separates management traffic from user data.
- Out-of-band has the management IP address in a separate subnet, VLAN, and switch for higher security.
- Out-of-band provides higher assurance for management data delivery during network failures.
- Out-of-band has no active Spanning Tree in management VLAN. Redundancy is not critical.

System Tests

Boot-up Diagnostics

During a system boot-up, a number of processes are performed in order to ensure that a reliable and operational platform is available so that faulty hardware does not disrupt the network. Catalyst boot diagnostics are split between Power-On Self Test (POST) and online diagnostics.

Operational Overview

Depending on the platform and hardware configuration, different diagnostics are carried out at boot-up and when a card is hot-swapped into the chassis. A higher level of diagnostics result in a wider number of problems detected but a longer boot cycle. These three levels of POST diagnostics can be selected (all tests check DRAM, RAM, and cache presence and size and initialize them):

Operational Overview			
Bypass	N/A	3	Not available on 4500/4000 series using CatOS 5.5 or earlier.
Minimal	Pattern-writing tests on the first MB of DRAM only.	30	Default on 5500/5000 and 6500/6000 series; not available on 4500/4000 series.
Complete	Pattern-writing tests for all memory.	60	Default on 4500/4000 series.

Online Diagnostics

These tests check packet paths internally in the switch. It is important to note that online diagnostics are therefore system-wide tests, not simply port tests. On Catalyst 5500/5000 and

6500/6000 switches, tests are performed first from the standby Supervisor Engine, and again from the primary Supervisor Engine. The length of the diagnostics depends on the system configuration (number of slots, modules, ports). There are three categories of tests:

- Loopback test—packets from the Supervisor Engine NMP are sent to each port, then returned to the NMP and examined for errors.
- Bundling test—channels of up to eight ports are created and loopback tests performed to the agport to verify the hashing to specific links (refer to the EtherChannel section of this document for further information).
- Enhanced Address Recognition Logic (EARL) test—both the central Supervisor Engine and in-line Ethernet module L3 rewrite engines are tested. Hardware forwarding entries and routed ports are created before sample packets are sent (for each protocol encapsulation type) from the NMP through the switching hardware on each module and back to the NMP. This is for Catalyst 6500/6000 PFC modules and newer.

Complete online diagnostics can take approximately two minutes. Minimal diagnostics do not perform bundle or rewrite testing on modules other then the Supervisor Engine, and can take approximately 90 seconds.

During a memory test, when a difference is found in the pattern read back compared to the pattern written, the port state is changed to `faulty`. The results of these tests can be seen if the **show test** command is issued, followed by the module number to be examined:

```
>show test 9

Diagnostic mode: complete (mode at next reset: complete)
!--- Configuration setting. Module 9 : 4-port Multilayer Switch Line Card Status for
Module 9 : PASS Port Status : Ports 1 2 3 4 ----------------- . . . . Line Card Diag
Status for Module 9 (. = Pass, F = Fail, N = N/A) Loopback Status [Reported by Module
1] : Ports 1 2 3 4 ----------------- . . F . !--- Faulty. Channel Status : Ports 1 2
3 4 ----------------- . . . .
```

Recommendation

Cisco recommends that all switches be set to use complete diagnostics to provide maximum fault detection and prevent outages during normal operations.

Note: This change does not take effect until the next time the device is booted. Issue this command in order to set complete diagnostics:

```
set test diaglevel complete
```

Other Options

In some situations, a rapid boot-up time can be preferable over waiting to run full diagnostics. There are other factors and timings involved in bringing up a system, but overall, POST and online diagnostics add around a third again in time. In testing with a fully populated single Supervisor Engine nine-slot chassis with a Catalyst 6509, the total boot time was around 380 seconds with complete diagnostics, around 300 seconds with minimal diagnostics, and only 250 seconds with

diagnostics bypassed. Issue this command to configure bypass:

```
set test diaglevel bypass
```

Note: The Catalyst 4500/4000 accepts being configured for minimal diagnostics, though this still results in a complete test being undertaken. Minimal mode could be supported in the future on this platform.

Run Time Diagnostics

Once the system is operational, the switch Supervisor Engine performs various monitoring of the other modules. If a module is not reachable through the management messages (Serial Control Protocol [SCP] running over the out-of-band management bus), the Supervisor Engine attempts to restart the card or take other action as appropriate.

Operational Overview

The Supervisor Engine carries out various monitoring automatically; this does not require any configuration. For the Catalyst 5500/5000 and 6500/6000, these components of the switch are monitored:

- NMP through a watchdog
- Enhanced EARL chip errors
- Inband channel from Supervisor Engine to backplane
- Modules through keepalives over out-of-band channel (Catalyst 6500/6000)
- Active Supervisor Engine is monitored by the standby Supervisor Engine for status (Catalyst 6500/6000)

System and Hardware Error Detection

Operational Overview

In CatOS 6.2 and later, further functionality has been added in order to monitor critical system and hardware-level components. These three hardware components are supported:

- Inband
- Port counter
- Memory

When the feature is enabled and an error condition is detected, the switch generates a syslog message. The message informs the administrator that a problem exists before noticeable performance degradation occurs. In CatOS versions 6.4(16), 7.6(12), 8.4(2) and later, the default mode for all three components changed from disabled to enabled.

Inband

If an inband error is detected, a syslog message informs you that a problem exists before noticeable performance degradation occurs. The error displays the type of inband failure

occurrence. Some examples are:

- Inband stuck
- Resource errors
- Inband fail during bootup

At the detection of an inband ping failure, the feature also reports an additional syslog message with a snapshot of the current Tx and Rx rate on the inband connection, CPU, and the backplane load of the switch. This message enables you to properly determine if the inband is stuck (no Tx/Rx) or overloaded (excessive Tx/Rx). This additional information can help you determine the cause of inband ping failures.

Port Counter

When you enable this feature, it creates and starts a process to debug port counters. The port counter periodically monitors select internal port error counters. The architecture of the line card, and more specifically the ASICs on the module, determines which counters the feature queries. Cisco Technical Support or development engineering can then use this information in order to troubleshoot problems. This feature does not poll error counters such as FCS, CRC, alignment, and runts that are directly associated with link partner connectivity. See the EtherChannel/Link Errors Handling section of this document in order to incorporate this capability.

Polling is executed every 30 minutes and runs in the background of selected error counters. If the count goes up between two subsequent polls on the same port, a syslog message reports the incident and gives the module/port and error counter details.

The port counter option is not supported on the Catalyst 4500/4000 platform.

Memory

Enablement of this feature performs background monitoring and detection of DRAM corruption conditions. Such memory corruption conditions include:

- Allocation
- Freeing
- Out of range
- Bad alignment

Recommendation

Enable all error detection features, which includes inband, port counters, and memory, where they are supported. Enablement of these features achieves improved proactive system and hardware warning diagnostics for the Catalyst switch platform. Issue these commands in order to enable all three error detection features:

```
set errordetection inband enable
!--- This is the default in CatOS 6.4(16), 7.6(12), 8.4(2), and later. set
errordetection portcounters enable
!--- This is the default in CatOS 6.4(16), 7.6(12), 8.4(2), and later. set
errordetection memory enable
```

```
!--- This is the default in CatOS 6.4(16), 7.6(12), 8.4(2), and later.
```

Issue this command in order to confirm that error detection is enabled:

```
>show errordetection

Inband error detection:                  enabled
Memory error detection:                  enabled
Packet buffer error detection:           errdisable
Port counter error detection:            enabled
Port link-errors detection:              disabled
Port link-errors action:                 port-failover
Port link-errors interval:               30 seconds
```

EtherChannel/Link Errors Handling

Operational Overview

In CatOS 8.4 and later, a new feature has been introduced in order to provide an automatic failover of traffic from one port in an EtherChannel to another port in the same EtherChannel. The port failover occurs when one of the ports in the channel exceeds a configurable error threshold within the specified interval. The port failover only occurs if there is an operational port left in the EtherChannel. If the failed port is the last port in the EtherChannel, the port does not enter the `port-failover` state. This port continues to pass traffic, regardless of the type of errors that are received. Single, nonchanneling ports do not go into the `port-failover` state. These ports go into the `errdisable` state when the error threshold is exceeded within the specified interval.

This feature is only effective when you enable **set errordetection portcounters**. The link errors to be monitored are based on three counters:

- InErrors
- RxCRCs (CRCAlignErrors)
- TxCRCs

Issue the **show counters** command on a switch in order to display the number of error counters. This is an example:

```
>show counters 4/48

.......

32 bit counters

0  rxCRCAlignErrors           =          0
.......

6  ifInErrors                 =          0
.......

12 txCRC                      =          0
```

This table is a list of possible configuration parameters and the respective default configuration:

Parameters	Default

Global	Disabled
Port monitor for RxCRC	Disabled
Port monitor for InErrors	Disabled
Port monitor for TxCRC	Disabled
Action	Port-failover
Interval	30 seconds
Sampling count	3 consecutive
Low threshold	1000
High threshold	1001

If the feature is enabled and the error count of a port reaches the high value of the configurable threshold within the specified sampling count period, the configurable action is either error disable or port failover. The error disable action places the port into the `errdisable` state. If you configure the port failover action, the port channel status is considered. The port is error disabled only if the port is in a channel but that port is not the last operational port in the channel. Additionally, if the configured action is port failover and the port is a single port or nonchanneled, the port is placed in the `errdisable` state when the port error count reaches the high value of the threshold.

The interval is a timer constant for reading the port error counters. The default value of the link-errors interval is 30 seconds. The allowed range is between 30 and 1800 seconds.

There is a risk of accidental error disablement of a port because of an unexpected one-time event. In order to minimize this risk, actions to a port are taken only when the condition persists through this consecutive sampling number of times. The default sampling value is 3 and the allowed range is from 1 to 255.

The threshold is an absolute number to be checked based on the link-errors interval. The default link-error low threshold is 1000 and the allowed range is 1 to 65,535. The default link-error high threshold is 1001. When the consecutive number of sampling times reaches the low threshold, a syslog is sent. If the consecutive sampling times reaches the high threshold, a syslog is sent and an error disable or port failover action is triggered.

Note: Use the same port error detection configuration for all ports in a channel. Refer to these sections of the Catalyst 6500 series software configuration guide for more information:

- The *Configuring EtherChannel/Link Error Handling* section of Checking Status and Connectivity
- The *Configuring Port Error Detection* section of Configuring Ethernet, Fast Ethernet, Gigabit Ethernet, and 10-Gigabit Ethernet Switching

Recommendations

Because the feature uses SCP messages in order to record and compare the data, high numbers of active ports can be CPU-intensive. This scenario is even more CPU-intensive when the threshold interval is set to a very small value. Enable this feature with discretion for ports that are designated as critical links and carry traffic for sensitive applications. Issue this command in order to enable link error detection globally:

```
set errordetection link-errors enable
```

Also, start with the default threshold, interval, and sampling parameters. And use the default action, port failover.

Issue these commands in order to apply the global link-error parameters to individual ports:

```
set port errordetection mod/port inerrors enable

set port errordetection mod/port rxcrc enable

set port errordetection mod/port txcrc enable
```

You can issue these commands in order to verify the link-errors configuration:

```
show errordetection

show port errordetection {mod | mod/port}
```

Catalyst 6500/6000 Packet Buffer Diagnostics

In CatOS versions 6.4(7), 7.6(5), and 8.2(1), the Catalyst 6500/6000 packet buffer diagnostics were introduced. The packet buffer diagnostics, which are enabled by default, detect packet buffer failures that are caused by transient Static RAM (SRAM) failures. Detection is on these 48-port 10/100-Mbps line modules:

- WS-X6248-RJ45
- WS-X6248-RJ21
- WS-X6348-RJ45
- WS-X6348-RJ21
- WS-X6148-RJ45
- WS-X6148-RJ21

When the failure condition occurs, 12 out of the 48 10/100-Mbps ports continue to stay connected and can experience random connectivity problems. The only way to recover from this condition is to power cycle the line module.

Operational Overview

The packet buffer diagnostics check the data that are stored in a specific section of the packet buffer in order to determine if it is corrupted by transient SRAM failures. If the process reads back something different than what it wrote, it then carries out two possible configurable recovery options:

1. The default action is to error disable the line card ports that are affected by the buffer failure.
2. The second option is to power cycle the line card.

Two syslog messages have been added. The messages provide a warning of the error disablement of the ports or the power cycle of the module because of packet buffer errors:

```
show errordetection

show port errordetection {mod | mod/port}
```

In CatOS versions that are earlier than 8.3 and 8.4, the line card power-cycle time is between 30 and 40 seconds. A Rapid Boot feature was introduced in CatOS versions 8.3 and 8.4. The feature automatically downloads the firmware to the installed line cards during the initial boot process in order to minimize the bootup time. The Rapid Boot feature reduces the power-cycle time to approximately 10 seconds.

Recommendation

Cisco recommends the default option of *errdisable*. This action has the least impact on the network service during production hours. If possible, move the connection that is affected by the error-disabled ports to other available switch ports in order to restore service. Schedule a manual power cycle of the line card during the maintenance window. Issue the **reset module *mod*** command in order to fully recover from the corrupted packet buffer condition.

Note: If the errors continue after the module is reset, try to re-seat the module.

Issue this command in order to enable the *errdisable* option:

```
set errordetection packet-buffer errdisable
!--- This is the default.
```

Other Option

Because a power cycle of the line card is necessary in order to fully recover all ports which have encountered an SRAM failure, an alternative recovery action is to configure the power cycle option. This option is useful in circumstances in which an outage in network services that can last between 30 and 40 seconds is acceptable. This length of time is the time that is necessary for a line module to fully power cycle and place itself back into service without the Rapid Boot feature. The Rapid Boot feature can reduce the time of the outage in network services to 10 seconds with the power cycle option. Issue this command in order to enable the power cycle option:

```
set errordetection packet-buffer power-cycle
```

Packet Buffer Diagnostics

This test is for Catalyst 5500/5000 switches only. This test is designed to find failed hardware on Catalyst 5500/5000 switches that are using Ethernet modules with specific hardware that provide 10/100-Mbps connectivity between user ports and the switch backplane. As they cannot perform

CRC checking for trunked frames, if a port packet buffer becomes defective during runtime, packets could get corrupted and cause CRC errors. Unfortunately, this could lead to the propagation of bad frames further into the Catalyst 5500/5000 ISL network, which potentially causes control plane disruption and broadcast storms in worst-case scenarios.

Newer Catalyst 5500/5000 modules and other platforms have updated hardware error checking built in and do not need the packet buffer tests, so there is no option to configure it.

The line modules that need the packet buffer diagnostics are WS-X5010, WS-X5011, WS-X5013, WS-X5020, WS-X5111, WS-X5113, WS-X5114, WS-X5201, WS-X5203, WS-X5213/a, WS-X5223, WS-X5224, WS-X5506, WS-X5509, WS-U5531, WS-U5533, and WS-U5535.

Operational Overview

This diagnostic checks that data stored in a specific section of the packet buffer is not accidentally being corrupted by faulty hardware. If the process reads back something different than it wrote, it shuts down the port in `failed` mode, since that port could corrupt data. There is no threshold of errors needed. Failed ports cannot be enabled again until the module has been reset (or replaced).

There are two modes for packet buffer tests: scheduled and on-demand. When a test begins, syslog messages are generated in order to indicate the expected length of the test (rounded up to the nearest minute) and the fact that the test has started. The exact length of the test varies by port type, size of the buffer, and the type of test run.

On-demand tests are aggressive in order to finish within a few minutes. Since these tests actively interfere with packet memory, ports must be administratively shut down before testing. Issue this command in order to shut down the ports:

```
> (enable) test packetbuffer 4/1
Warning: only disabled ports may be tested on demand - 4/1 will be skipped.
> (enable) set port disable 4/1
> (enable) test packetbuffer 4/1
Packet buffer test started. Estimated test time: 1 minute.
%SYS-5-PKTTESTSTART:Packet buffer test started
%SYS-5-PKTTESTDONE:Packet buffer test done. Use 'show test' to see test results
```

Scheduled tests are much less aggressive than the on-demand tests, and they execute in the background. The tests are performed in parallel across multiple modules but on one port per module at a time. The test preserves, writes, and reads small sections of packet buffer memory before restoring user packet buffer data, and thus generates no errors. However, since the test is writes to buffer memory, it blocks incoming packets for a few milliseconds and causes some loss on busy links. By default there is an eight-second pause between each buffer-write test to minimize any packet loss, but this means that a system full of modules that need the packet buffer test can take over 24 hours for the test to complete. This scheduled test is enabled by default to run weekly at 03:30 on Sundays from CatOS 5.4 or later, and the test status can be confirmed with this command:

```
>show test packetbuffer status

!--- When test is running, the command returns !--- this information: Current packet
buffer test details Test Type : scheduled Test Started : 03:30:08 Jul 20 2001 Test
```

```
Status : 26% of ports tested Ports under test : 10/5,11/2 Estimated time left : 11
minutes !--- When test is not running, !--- the command returns this information:
Last packet buffer test details Test Type : scheduled Test Started : 03:30:08 Jul 20
2001 Test Finished : 06:48:57 Jul 21 2001
```

Recommendation

The Cisco recommends that you use the scheduled packet buffer test feature for Catalyst 5500/5000 systems, as the benefit of discovering problems on modules outweighs the risk of low packet loss.

A standardized weekly time must then be scheduled across the network that allows the customer to change links from faulty ports or RMA modules as necessary. As this test can cause some packet loss, depending on network load, it must be scheduled for quieter network times, such as the default of 3:30 AM on a Sunday morning. Issue this command in order to set the test time:

```
set test packetbuffer Sunday 3:30
!--- This is the default.
```

Once enabled (as when CatOS is upgraded to 5.4 and later for the first time), there is a chance that a previously hidden memory/hardware problem is exposed, and a port is shut down automatically as a result. You could see this message:

```
set test packetbuffer Sunday 3:30
!--- This is the default.
```

Other Options

If it is not acceptable to risk a low level of packet loss per-port on a weekly basis, then it is recommended to use the on-demand feature during scheduled outages. Issue this command in order to start this feature manually on a per range basis (though the port must be administratively disabled first):

```
test packetbuffer port range
```

System Logging

Syslog messages are Cisco-specific and a key part of proactive fault management. A wider range of network and protocol conditions are reported using syslog than is possible through standardized SNMP. Management platforms, such as Cisco Resource Manager Essentials (RMEs) and the Network Analysis Toolkit (NATkit) make powerful use of syslog information because they perform these tasks:

- Present analysis by severity, message, device, and so on
- Enable filtering of messages coming in for analysis
- Trigger alerting, such as pagers, or on-demand collecting of inventory and configuration

changes

Recommendation

An important point of focus is what level of logging information is to be generated locally and held in the switch buffer as opposed to that which is sent to a syslog server (using the **set logging server severity *value*** command). Some organizations log a high level of information centrally, whereas others go to the switch itself to look at the more detailed logs for an event or enable a higher level of syslog capture only during troubleshooting.

Debugging is different on CatOS platforms than Cisco IOS Software, but detailed system logging can be enabled on a per-session basis with **set logging session enable** without changing what is logged by default.

Cisco generally recommends that you bring the spantree and system syslog facilities up to level 6, as these are key stability features to track. In addition, for multicast environments, bringing the logging level of the mcast facility up to 4 is recommended so that syslog messages are produced if router ports are deleted. Unfortunately, before CatOS 5.5(5) this could result in syslog messages being recorded for IGMP joins and leaves, which is too noisy to monitor. Finally, if IP input lists are used, a minimum logging level of 4 is recommended to capture unauthorized login attempts. Issue these commands in order to set these options:

```
set logging buffer 500
!--- This is the default. set logging server syslog server IP address set logging
server enable
!--- This is the default. set logging timestamp enable
set logging level spantree 6 default
!--- Increase default STP syslog level. set logging level sys 6 default
!--- Increase default system syslog level. set logging server severity 4
!--- This is the default; !--- it limits messages exported to syslog server. set
logging console disable
```

Turn off the console messages in order to protect against the risk of the switch hanging as it waits for a response from a slow or non-existing terminal when message volume is high. Console logging is a high priority under CatOS and is mainly used to capture the final messages locally when troubleshooting or in a switch crash scenario.

This table provides the individual logging facilities, default levels, and recommended changes for the Catalyst 6500/6000. Each platform has slightly different facilities, depending on the features supported.

Facility	Default Level	Recommended Action
acl	5	Leave alone.
cdp	4	Leave alone.
cops	3	Leave alone.
dtp	8	Leave alone.
earl	2	Leave alone.
ethc[1]	5	Leave alone.

filesys	2	Leave alone.
gvrp	2	Leave alone.
ip	2	**Change to 4 if IP input lists used**.
kernel	2	Leave alone.
1d	3	Leave alone.
mcast	2	**Change to 4 if multicast used (CatOS 5.5[5] and later)** .
mgmt	5	Leave alone.
mls	5	Leave alone.
pagp	5	Leave alone.
protfilt	2	Leave alone.
pruning	2	Leave alone.
Privatevlan	3	Leave alone.
qos	3	Leave alone.
radius	2	Leave alone.
rsvp	3	Leave alone.
security	2	Leave alone.
snmp	2	Leave alone.
spantree	2	**Change to 6**.
sys	5	**Change to 6**.
tac	2	Leave alone.
tcp	2	Leave alone.
telnet	2	Leave alone.
Tftp	2	Leave alone.
UDLD	4	Leave alone.
VMPS	2	Leave alone.
VTP	2	Leave alone.

[1] In CatOS 7.x and later, the ethc facility code replaces the pagp facility code in order to reflect LACP support.

Note: Currently, the Catalyst switches log a configuration change syslog level-6 message for each **set** or **clear** command executed, unlike Cisco IOS Software, which triggers the message only after you exit configuration mode. If you need RMEs to back up configurations in real-time upon this trigger, then these messages also need to be sent to the RMEs syslog server. For most customers, periodic configuration backups for Catalyst switches are enough, and no change of the default server logging severity is needed.

If you tune your NMS alerts, consult the System Message Guide.

Simple Network Management Protocol

SNMP is used to retrieve statistics, counters, and tables stored in network device Management Information Bases (MIBs). The information collected can be used by NMSs (such as HP Openview) in order to generate real time alerts, measure availability, and produce capacity

planning information, as well as to help perform configuration and troubleshooting checks.

Operational Overview

With some security mechanisms, a network management station is able to retrieve information in the MIBs with SNMP protocol get and get next requests, and to change parameters with the **set** command. Additionally, a network device can be configured to generate a trap message for the NMS for real-time alerting. SNMP polling uses IP UDP port 161 and SNMP traps use port 162.

Cisco supports these versions of SNMP:

- SNMPv1: RFC 1157 Internet Standard, using clear text community string security. An IP address access control list and password define the community of managers able to access the agent MIB.
- SNMPv2C: a combination of SNMPv2, a draft Internet standard defined in RFCs 1902 through 1907, and SNMPv2C, a community-based administrative framework for SNMPv2 that is an experimental draft defined in RFC 1901. Benefits include a Bulk retrieval mechanism that supports the retrieval of tables and large quantities of information, minimizes the number of round-trips required, and improves error handling.
- SNMPv3: RFC 2570 proposed draft provides secure access to devices through the combination of authentication and encryption of packets over the network. The security features provided in SNMPv3 are:Message integrity: ensures that a packet has not been tampered with in-transitAuthentication: determines that the message is from a valid sourceEncryption: scrambles the contents of a packet to prevent it from being viewed easily by an unauthorized source

This table identifies the combinations of security models:

Model Level	**Authentication**	**Encryption**	**Result**
v1	noAuthNoPriv, Community String	No	Uses a community string match for authentication.
v2c	noAuthNoPriv, Community String	No	Uses a community string match for authentication.
v3	noAuthNoPriv, Username	No	Uses a username match for authentication.
v3	authNoPriv, MD5 or SHA	Np	Provides authentication based on the HMAC-MD5 or HMAC-SHA algorithms.

v3	authPriv, MD5 or SHA	DES	Provides authentication based on the HMAC-MD5 or HMAC-SHA algorithms. Provides DES 56-bit encryption in addition to authentication based on the CBC-DES (DES-56) standard.

Note: Keep this information in mind about SNMPv3 objects:

- Each user belongs to a group.
- A group defines the access policy for a set of users.
- An access policy defines what SNMP objects can be accessed to read, write, and create.
- A group determines the list of notifications its users can receive.
- A group also defines the security model and security level for its users.

SNMP Trap Recommendation

SNMP is the foundation of all network management and is enabled and used on all networks. The SNMP agent on the switch must be set to use the version of SNMP supported by the management station. Since an agent can communicate with multiple managers, it is possible to configure the software to support communication with one management station using the SNMPv1 protocol and another using the SNMPv2 protocol, for example.

Most NMS stations use SNMPv2C today under this configuration:

```
set snmp community read-only string
!--- Allow viewing of variables only. set snmp community read-write string
!--- Allow setting of variables. set snmp community read-write-all string<string>
!--- Include setting of SNMP strings.
```

Cisco recommends that SNMP traps be enabled for all features in use (features not used can be disabled if desired). Once a trap is enabled, it can be tested with the **test snmp** command and appropriate handling set up on the NMS for the error (such as a pager alert or pop-up).

All traps are disabled by default and need to be added to the configuration, either individually or by with the **all** parameter, as shown:

```
set snmp trap enable all
set snmp trap server address read-only community string
```

Available traps in CatOS 5.5 include:

Trap	Description
auth	Authentication
bridge	Bridge
chassis	Chassis

config	Configuration
entity	Entity
ippermit	IP permit
module	Module
repeater	Repeater
stpx	Spanning Tree extension
syslog	Syslog notification
vmps	VLAN Membership Policy Server
vtp	VLAN Trunk Protocol

Note: The syslog trap sends all syslog messaged generated by the switch to the NMS as a SNMP trap also. If syslog alerting is already being performed by an analyzer such as Cisco Works 2000 RMEs, then it is not necessarily useful to receive this information twice.

Unlike Cisco IOS Software, port level SNMP traps are disabled by default because switches can have hundreds of active interfaces. Cisco therefore recommends that key ports, such as infrastructure links to routers, switches, and main servers, have port-level SNMP traps enabled. Other ports, like user host ports, are not required, which helps to simplify network management.

```
set port trap port range enable
!--- Enable on key ports only.
```

SNMP Polling Recommendation

A network management review is recommended in order to discuss specific needs in detail. However, some basic Cisco philosophies for the management of large networks are listed:

- Do something simple, and do it well.
- Reduce staff overload due to excessive data polling, collection, tools, and manual analysis.
- Network management is possible with just a few tools, such as HP Openview as an NMS, Cisco RMEs as a configuration, syslog, inventory, and software manager, Microsoft Excel as an NMS data analyzer, and CGI as a way to publish to the web.
- Publishing reports to the web allows users, such as senior management and analysts, to help themselves to information without burdening operations staff with many special requests.
- Find out what is working well on the network and leave it alone. Concentrate on what is not working.

The first phase of NMS implementation must be to baseline the network hardware. Much can be inferred about device and protocol health from simple CPU, memory, and buffer utilization on routers, and NMP CPU, memory, and backplane utilization on switches. Only after a hardware baseline do L2 and L3 traffic load, peak, and average baselines become fully meaningful. Baselines are usually established over several months to get visibility of daily, weekly, and quarterly trends – according to the business cycle of the company.

Many networks suffer NMS performance and capacity problems caused by over-polling. It is therefore recommended, once the baseline is established, to set alarm and event RMON thresholds on the devices themselves to alert the NMS on abnormal changes, and thus remove polling. This enables the network to tell the operators when something is not normal rather than continuously polling to see whether everything is normal. Thresholds can be set based on various

rules, such as maximum value plus a percentage or standard deviation from a mean, and are outside the scope of this document.

The second phase of NMS implementation is to poll particular areas of the network in more detail with SNMP. This includes areas of doubt, areas before a change, or areas that are be characterized as working well. Use the NMS systems as a searchlight to scan the network in detail and illuminate hot spots (do not attempt to light up the whole network).

The Cisco Network Management Consulting group suggests these key fault MIBs to be analyzed or monitored in campus networks. Refer to Cisco Network Monitoring and Event Correlation Guidelines for more information (on performance MIBs to poll, for example).

Object Name	Object Description	OID	Poll Interval	Threshold
MIB-II				
sysUpTime	system uptime in 1/100ths of seconds	1.3.6.1.2.1.1.3	5 min	< 30000

Object Name	Object Description	OID	Poll Interval	Threshold
CISCO-PROCESS-MIB				
cpmCPUTotal5min	The overall CPU busy percentage in the last 5 minute period	1.3.6.1.4.1.9.9.109.1.1.1.1.5	10 min	Baseline

Object Name	Object Description	OID	Poll Interval	Threshold
CISCO-STACK-MIB				
sysEnableChassisTraps	Indicates whether chassisAlarmOn and chassisAlarmOff traps in this MIB must be generated.	1.3.6.1.4.1.9.5.1.1.24	24 hrs	1
sysEnableModuleTraps	Indicates whether	1.3.6.1.4.1.9.5.1.1.25	24 hrs	1

	moduleUp and moduleDown traps in this MIB must be generated.			
sysEnableBridgeTraps	Indicates whether newRoot and topologyChange traps in the BRIDGE-MIB (RFC 1493) must be generated.	1.3.6.1.4.1.9.5.1.1.26	24 hrs	1
sysEnableRepeaterTraps	Indicates whether the traps in the REPEATER-MIB (RFC1516) must be generated.	1.3.6.1.4.1.9.5.1.1.29	24 hrs	1
sysEnableIpPermitTraps	Indicates whether the IP permit traps in this MIB must be generated.	1.3.6.1.4.1.9.5.1.1.31	24 hrs	1
sysEnableVmpsTraps	Indicates whether the vmVmpsChange trap defined in CISCO- VLAN-MEMBERSHIP-MIB must be generated.	1.3.6.1.4.1.9.5.1.1.33	24 hrs	1
sysEnableConfigTraps	Indicates whether sysConfigChange trap in this MIB must be generated.	1.3.6.1.4.1.9.5.1.1.35	24 hrs	1
sysEnableStpxTrap	Indicates whether stpxInconsistencyUpdate trap in the CISCO-STP-EXTENSIONS-MIB must be generated.	1.3.6.1.4.1.9.5.1.1.40	24 hrs	1
chassisPs1status	Status of power supply 1.	1.3.6.1.4.1.9.5.1.2.4	10 min	2
chassisPs1TestResult	Detailed information on	1.3.6.1.4.1.9.5.1.2.5	As nee	

	status of power supply 1.		ded.	
chassisPs2Status	Status of power supply 2.	1.3.6.1.4.1.9.5.1.2.7	10 min	2
chassisPs2TestResult	Detailed information on status of power supply 2	1.3.6.1.4.1.9.5.1.2.8	As needed.	
chassisFanStatus	Status of Chassis Fan.	1.3.6.1.4.1.9.5.1.2.9	10 min	2
chassisFanTestResult	Detailed information on status of chassis fan.	1.3.6.1.4.1.9.5.1.2.10	As needed.	
chassisMinorAlarm	Chassis Minor Alarm Status.	1.3.6.1.4.1.9.5.1.2.11	10 min	1
chassis MajorAlarm	Chassis Major Alarm Status	1.3.6.1.4.1.9.5.1.2.12	10 min	1
chassisTempAlarm	Chassis Temperature Alarm status.	1.3.6.1.4.1.9.5.1.2.13	10 min	1
moduleStatus	Operational Status of the module.	1.3.6.1.4.1.9.5.1.3.1.1.10	30 min	2
moduleTest Result	Detailed information on modules condition.	1.3.6.1.4.1.9.5.7.3.1.1.11	As needed.	
moduleStandbyStatus	Status of a redundant module.	1.3.6.1.4.1.9.5.7.3.1.1.21	30 min	=1 or =4

Object Name	**Object Description**	**OID**	**Poll Interval**	**Threshold**
CISCO-MEMORY-POOL-MIB				
dot1dStpTimeSinceTopologyChange	The time (in 1/100 secs) since the last time a topology change was detected by the entity.	1.3.6.1.2.1.17.2.3	5 min	< 30000
dot1dStpTopChanges	The total number of topology changes	1.3.6.1.2.1.17.2.4	As needed.	

	detected by this bridge since the managemen t entity was last reset or initialized.			
dot1dStpPortState [1]	The current state of the port as defined by application of the Spanning Tree Protocol. Return value can be one of these: `disabled (1), blocking (2), listening (3), learning (4), forwarding (5),` or `broken (6).`	1.3.6.1.2.1. 17.2.15.1.3	As nee ded .	

Object Name	**Object Description**	**OID**	**Pol l Int erv al**	**Thre shol d**
CISCO-MEMORY-POOL-MIB				
ciscoMemoryP oolUsed	Indicates the number of bytes from the memory pool that are currently in use by applications on the managed device.	1.3.6.1.4.1.9 .9.48.1.1.1.5	30 min	Base line
ciscoMemoryP oolFree	Indicates the number of bytes from the memory pool that are currently	1.3.6.1.4.1.9 .9.48.1.1.1.6	30 min	Base line

	unused on the managed device. **Note:** The sum of ciscoMemoryPoolUsed and ciscoMemoryPoolFree is the total amount of memory in the pool.			
ciscoMemoryPoolLargestFree	Indicates the largest number of contiguous bytes from the memory pool that are currently unused on the managed device.	1.3.6.1.4.1.9.9.48.1.1.1.7	30 min	Baseline

Refer to Cisco Network Management Toolkit - MIBs for more information on Cisco MIB support.

Note: Some standard MIBs assume that a particular SNMP entity contains only one instance of the MIB. Thus, the standard MIB does not have any index that allows users to directly access a particular instance of the MIB. In these cases, community string indexing is provided in order to access each instance of the standard MIB. The syntax is [community string]@[instance number], where instance is typically a VLAN number.

Other Options

The security aspects of SNMPv3 mean that its use is expected to overtake SNMPv2 in time. Cisco recommends that customers prepare for this new protocol as part of their NMS strategy. The benefits are that data can be collected securely from SNMP devices without fear of tampering or corruption. Confidential information, such as SNMP **set** command packets that change a switch configuration, can be encrypted to prevent its contents from being exposed on the network. In addition, different user groups can have different privileges.

Note: The configuration of SNMPv3 is significantly different than the SNMPv2 command line, and increased CPU load on the Supervisor Engine is to be expected.

Remote Monitoring

RMON permits MIB data to be pre-processed by the network device itself, in preparation for common uses or application of that information by the network manager, such as performing historical baseline determination and threshold analysis.

The results of RMON processing are stored in RMON MIBs for subsequent collection by an NMS,

as defined in RFC 1757 .

Operational Overview

Catalyst switches support mini-RMON in hardware on each port, which consists of four basic RMON-1 groups: Statistics (group 1), History (group 2), Alarms (group 3), and Events (group 9).

The most powerful part of RMON-1 is the **threshold mechanism** provided by the **alarm and event** groups. As discussed, the configuration of RMON thresholds allows the switch to send an SNMP trap when an anomalous condition occurs. Once key ports have been identified, SNMP can be used in order to poll counters or RMON history groups and create baselines recording normal traffic activity for those ports. Next, RMON rising and falling thresholds can be set and alarms configured for when there is a defined variance from the baseline.

Configuration of thresholds is best performed with an RMON management package, since the successful creation of the rows of parameters in Alarm and Event tables is tedious. Commercial RMON NMS packages, such as the Cisco Traffic Director, part of Cisco Works 2000, incorporate GUIs that make the setting of RMON thresholds much simpler.

For baseline purposes, the etherStats group provides a useful range of L2 traffic statistics. The objects in this table can be used to get statistics on unicast, multicast, and broadcast traffic as well as a variety of L2 errors. The RMON agent on the switch can also be configured to store these sampled values in the history group. This mechanism enables the amount of polling to be reduced without reducing the sample rate. RMON histories can give accurate baselines without substantial polling overhead. However, the more histories collected, the more switch resources are used.

While switches provide only four basic groups of RMON-1, it is important not to forget the rest of RMON-1 and RMON-2. All groups are defined in RFC 2021, including UsrHistory (group 18) and ProbeConfig (group 19). L3 and higher information can be retrieved from switches with the SPAN port or VLAN ACL redirect features that enable you to copy traffic to an external RMON SwitchProbe or an internal Network Analysis Module (NAM).

NAMs support all RMON groups and can even examine **application layer data**, including Netflow data exported from Catalysts when MLS is enabled. Running MLS means that the router does not switch all packets in a flow, so only Netflow data-export and not interface counters give reliable VLAN accounting.

You can use a SPAN port and a switch probe to capture a packet stream for a particular port, trunk, or VLAN and upload the packets to decode with a RMON management package. The SPAN port is SNMP-controllable through the SPAN group in the CISCO-STACK-MIB, so this process is easy to automate. The Traffic Director makes use of these features with its roving agent feature.

There are caveats to spanning a whole VLAN. Even if you use a 1Gbps probe, the entire packet stream from one VLAN or even one 1Gbps full-duplex port can exceed the bandwidth of the SPAN port. If the SPAN port is continuously running at full bandwidth, chances are data is being lost. Refer to Configuring the Catalyst Switched Port Analyzer (SPAN) Feature for more details.

Recommendation

Cisco recommends that RMON thresholds and alerting be deployed in order to help network management in a more intelligent way than SNMP polling alone. This reduces network management traffic overhead and allows the network to alert intelligently when something has

changed from the baseline. RMON needs to be driven by an external agent such as Traffic Director; there is no CLI support. Issue these commands in order to enable RMON:

```
set snmp rmon enable
set snmp extendedrmon netflow enable mod
!--- For use with NAM module only.
```

It is important to remember that the primary function of a switch is to forward frames, not to act as a large multi-port RMON probe. Therefore, as you set up histories and thresholds on multiple ports for multiple conditions, keep in mind that resources are being consumed. Consider a NAM module if you are scaling up RMON. Also remember the critical port rule: only poll and set thresholds on the ports identified as important in the planning stage.

Memory Requirements

RMON memory usage is constant across all switch platforms relating to statistics, histories, alarms, and events. RMON uses a bucket in order to store histories and statistics on the RMON agent (the switch, in this case). The bucket size is defined on the RMON probe (Switch Probe) or RMON application (Traffic Director), then sent to the switch in order to be set. Typically, memory constraints are only a consideration on older Supervisor Engines with less than 32MB of DRAM. Refer to these guidelines:

- Approximately 450K of code space is added to the NMP image in order to support mini-RMON (which is four groups of RMON: statistics, history, alarms, and events). The dynamic memory requirement for RMON varies because it depends on the run-time configuration. The run-time RMON memory usage information for each mini-RMON group is explained here:Ethernet Statistics group—Takes 800 bytes for each switched Ethernet/FE interface.History group—For the Ethernet interface, each configured history control entry with 50 buckets takes approximately 3.6KB memory space and 56 bytes for each additional bucket.Alarms and Events groups—Takes 2.6KB for each configured alarm and its corresponding event entries.
- To save the RMON-related configuration takes approximately 20K NVRAM of space if the system total NVRAM size is 256K or more and 10K NVRAM of space if the total NVRAM size is 128K.

Network Time Protocol

The NTP, RFC 1305 , synchronizes timekeeping among a set of distributed time-servers and clients and allows events to be correlated when system logs are created or other time-specific events occur.

NTP provides client time accuracies, typically within a millisecond on LANs and up to a few tens of milliseconds on WANs, relative to a primary server synchronized to Coordinated Universal Time (UTC). Typical NTP configurations utilize multiple redundant servers and diverse network paths in order to achieve high accuracy and reliability. Some configurations include cryptographic authentication in order to prevent accidental or malicious protocol attacks.

Operational Overview

NTP was first documented in RFC 958 , but has evolved through RFC 1119 (NTP version 2) and

is now in its third version as defined in RFC 1305 . It runs over the UDP port 123. All NTP communication uses UTC, which is the same time as Greenwich Mean Time.

Accessing Public Time Servers

The NTP subnet presently includes over 50 public primary servers synchronized directly to UTC by radio, satellite, or modem. Normally, client workstations and servers with a relatively small number of clients do not synchronize to primary servers. There are about 100 public secondary servers synchronized to the primary servers that provide synchronization to over 100,000 clients and servers on the Internet. The current lists are maintained on the List of Public NTP Servers page, which is updated regularly. There are numerous private primary and secondary servers not normally available to the public as well. For a list of public NTP servers and information about how to use them, consult the University of Delaware Time Synchronization Server website.

Since there is no guarantee that these public Internet NTP servers will be available, or that they produce the correct time, it is strongly advised that other options be considered. This could include the use of various standalone Global Positioning Service (GPS) devices directly connected to a number of routers.

Another possible option is the use of various routers configured as Stratum 1 masters, although this is not recommended.

Stratum

Each NTP server adopts a stratum that indicates how far away from an external source of time the server is. Stratum 1 servers have access to some kind of external time source, such as a radio clock. Stratum 2 servers obtain time details from a nominated set of Stratum 1 servers, while Stratum 3 servers obtain time details from Stratum 2 servers, and so on.

Server Peer Relationship

- A server is one that responds to client requests, but does not try to incorporate any date information from a client time source.
- A peer is one that responds to client requests, but tries to use the client requests as being a potential candidate for a better time source and to aid in stabilization of its clock frequency.
- In order to be a true peer, both sides of the connection must enter into a peer relationship rather than have one user a peer and the other user a server. It is also recommended that peers exchange keys so that only trusted hosts talk to each other as peers.
- In a client request to a server, the server answers the client and forgets that the client ever asked a question; in a client request to a peer, the server answers the client and keeps state information about the client to track how well it is doing at timekeeping and what stratum server it is running.**Note:** CatOS can only act as an NTP client.

It is no problem for an NTP server to handle many thousands of clients. However, handling hundreds of peers has a memory impact, and the state maintenance consumes more CPU resources on the box as well as bandwidth.

Polling

The NTP protocol allows a client to query a server any time it wishes. In fact, when NTP is first

configured in a Cisco device, it sends out eight queries in rapid succession at NTP_MINPOLL (24 = 16 second) intervals. The NTP_MAXPOLL is 214 seconds (which is 16,384 seconds or 4 hours, 33 minutes, 4 seconds), the maximum time it takes before NTP polls again for a response. At present, Cisco does not have a method to manually force the POLL time to be set by the user.

The NTP polling counter starts at 2^6 (64) seconds and is incremented by powers of two (as the two servers sync with each other), to 2^{10}. That is, you can expect the sync messages to be sent at an interval of 64, 128, 256, 512, or 1024 seconds per configured server or peer. The time varies between 64 seconds and 1024 seconds as a power of two based on the phase-locked-loop that sends and receives packets. If there is a lot of jitter in the time, it polls more often. If the reference clock is accurate and the network connectivity consistent, you see the poll-times converge on 1024 seconds between each poll.

In the real world, this means that the NTP Poll Interval changes as the connection between the client and server changes. The better the connection, the longer the poll interval, meaning that the NTP client has received eight responses for its last eight requests (the poll interval is then be doubled). A single missed response causes the poll interval to be halved. The poll interval starts out at 64 seconds and goes to a maximum of 1024 seconds. In the best circumstances, it takes a little over two hours for the poll interval to go from 64 seconds to 1024 seconds.

Broadcasts

NTP broadcasts are never forwarded. The **ntp broadcast** command causes the router to originate NTP broadcasts on the interface on which it is configured. The **ntp broadcastclient** command causes the router or switch to listen to NTP broadcasts on the interface on which it is configured.

NTP Traffic Levels

The bandwidth utilized by NTP is minimal, since the interval between polling messages exchanged between peers usually ratchets back to no more than one message every 17 minutes (1024 seconds). With careful planning, this can be maintained within router networks over the WAN links. The NTP clients must peer to local NTP servers, not all the way across the WAN to the central site core routers who will be the stratum 2 servers.

A converged NTP client uses approximately 0.6 bits/second per server.

Recommendation

Many customers have NTP configured in client mode today on their CatOS platforms, synchronized from several reliable feeds from the Internet or a radio clock. However, a simpler alternative to server mode when operating a large number of switches is to enable NTP in broadcast client mode on the management VLAN in a switched domain. This mechanism allows an entire domain of Catalysts to receive a clock from a single broadcast message. However, the accuracy of timekeeping is marginally reduced because the information flow is one-way.

Using loopback addresses as the source of updates can also help with consistency. Security concerns can be addressed in these two ways:

- Filtering server updates
- Authentication

Time correlation of events is extremely valuable in two cases: troubleshooting and security audits.

Care must be taken in order to protect the time sources and data, and encryption is recommended so that key events are not erased either intentionally or unintentionally.

Cisco recommends these configurations:

Catalyst Configuration

```
set ntp broadcastclient enable
set ntp authentication enable
set ntp key key
!--- This is a Message Digest 5 (MD5) hash. set ntp
timezone <zone name>
set ntp summertime <date change details>
```

Alternate Catalyst Configuration

```
!--- This more traditional configuration creates !---
more configuration work and NTP peerings. set ntp client
enable
set ntp server IP address of time server set timezone
zone name set summertime date change details
```

Router configuration

```
!--- This is a sample router configuration to distribute
!--- NTP broadcast information to the Catalyst broadcast
clients. ntp source loopback0
ntp server IP address of time server ntp update-calendar
clock timezone zone name clock summer-time date change
details ntp authentication key key ntp access-group
access-list
!--- To filter updates to allow only trusted sources of
NTP information. Interface to campus/management VLAN
containing switch sc0 ntp broadcast
```

Cisco Discovery Protocol

CDP exchanges information between adjacent devices over the data link layer and is extremely helpful in the determination of the network topology and physical configuration outside of the logical or IP layer. Supported devices are mainly switches, routers, and IP phones. This section highlights some of the enhancements of CDP version 2 over version 1.

Operational Overview

CDP uses SNAP encapsulation with type code 2000. On Ethernet, ATM, and FDDI, the destination multicast address **01-00-0c-cc-cc-cc, HDLC protocol type 0x2000** is used. On Token Rings, the functional address c000.0800.0000 is used. CDP frames are sent periodically every minute by default.

CDP messages contain one or more sub-messages that allow the destination devices to gather and store information about every neighbor device.

CDP version 1 supports these parameters:

Parameter	Type	Description
1	Device-ID	Hostname of the device or hardware serial number in ASCII.
2	Address	The L3 address of the interface that has sent the update.
3	Port-ID	The port on which the CDP update has been sent.
4	Capabilities	Describes the functional capabilities of the device: Router: 0x01 TB Bridge: 0x02 SR Bridge: 0x04 Switch: 0x08 (Provides L2 and/or L3 switching) Host: 0x10 IGMP conditional filtering: 0x20 The Bridge or Switch does not forward IGMP report packets on non-routerports. Repeater: 0x40
5	Version	A character string containing the software version (same as in **show version**).
6	Platform	Hardware platform, such as WS-C5000, WS-C6009, or Cisco RSP.

In CDP version 2, additional protocol fields have been introduced. CDP version 2 supports any field, but the ones listed can be particularly useful in switched environments and are used in CatOS.

Note: When a switch runs CDPv1, it drops v2 frames. When a switch running CDPv2 receives a CDPv1 frame on an interface, it starts to send out CDPv1 frames out of that interface in addition to CDPv2 frames.

Parameter	Type	Description
9	VTP Domain	The VTP Domain, if configured on the device.
10	Native VLAN	In dot1q, this is the untagged VLAN.
11	Full/Half Duplex	This field contains the duplex setting of the sending port.

Recommendation

CDP is enabled by default and is essential to gain visibility of adjacent devices and for troubleshooting. It is also used by network management applications to build L2 topology maps. Issue these commands in order to set up CDP:

```
set cdp enable
!--- This is the default. set cdp version v2
!--- This is the default.
```

In parts of the network where a high level of security is required (such as Internet-facing DMZs),

CDP must be turned off as such:

```
set cdp disable port range
```

The **show cdp neighbors** command displays the local CDP table. Entries marked with a star (*) indicate a VLAN mismatch; entries marked with a # indicate a duplex mismatch. This can be a valuable help for troubleshooting.

```
>show cdp neighbors

* - indicates vlan mismatch.
# - indicates duplex mismatch.
Port  Device-ID          Port-ID Platform
----- ------------------ ------- ------------
 3/1  TBA04060103(swi-2) 3/1     WS-C6506
 3/8  TBA03300081(swi-3) 1/1     WS-C6506
15/1  rtr-1-msfc         VLAN 1  cisco    Cat6k-MSFC
16/1  MSFC1b             Vlan2   cisco    Cat6k-MSFC
```

Other Options

Some switches, like the Catalyst 6500/6000, have the ability to supply power by way of UTP cables to IP phones. Information received by way of CDP assists power management on the switch.

As IP phones can have a PC connected to them, and both devices connect to the same port on the Catalyst, the switch has the ability to put the VoIP phone in a separate VLAN, the auxiliary. This allows the switch to easily apply a different Quality of Service (QoS) for the VoIP traffic.

In addition, if the auxiliary VLAN is modified (for example, in order to force the phone to use a specific VLAN or specific tagging method), this information is sent to the phone by way of CDP.

Parameter	Type	Description
14	Appliance ID	Allows the VoIP traffic to be differentiated from other traffic, as by separate VLAN-id (auxiliary VLAN).
16	Power Consumption	The amount of power a VoIP phone consumes, in milliwatts.

Note: Catalyst 2900 and 3500XL switches do not currently support CDPv2.

Security Configuration

Ideally, the customer has already established a security policy to help define what tools and technologies from Cisco are qualified.

Note: Cisco IOS Software security, as opposed to CatOS, is dealt with in many documents, such as Cisco ISP Essentials.

Basic Security Features

Passwords

Configure a user level password (login). Passwords are case sensitive in CatOS 5.x and later, and can be from 0 to 30 characters in length, including spaces. Set the enable password:

```
set password password set enablepass password
```

All passwords must meet minimum length standards (for example, six characters minimum, a mix of letters and numbers, upper- and lower-case letters) for login and enable passwords when used. These passwords are encrypted using the MD5 hashing algorithm.

In order to allow for more flexibility in managing password security and device access, Cisco recommends the use of a TACACS+ server. Refer to the TACACS+ section of this document for more information.

Secure Shell

Utilize SSH encryption in order to provide security for Telnet sessions and other remote connections to the switch. SSH encryption is supported for remote logins to the switch only. You cannot encrypt Telnet sessions that are initiated from the switch. SSH version 1 is supported in CatOS 6.1, and version 2 support was added in CatOS 8.3. SSH version 1 supports the Data Encryption Standard (DES) and Triple-DES (3-DES) encryption methods, and SSH version 2 supports the 3-DES and Advanced Encryption Standard (AES) encryption methods. You can use SSH encryption with RADIUS and TACACS+ authentication. This feature is supported with SSH (k9) images. Refer to How to Configure SSH on Catalyst Switches Running CatOS for details.

```
set crypto key rsa 1024
```

In order to disable version 1 fallback and accept version 2 connections, issue this command:

```
set ssh mode v2
```

IP Permit Filters

These are filters to safeguard access to the management sc0 interface through Telnet and other protocols. These are particularly important when the VLAN used for management also contains users. Issue these commands in order to enable IP address and port filtering:

```
set ip permit enable
set ip permit IP address mask Telnet|ssh|snmp|all
```

However, if Telnet access is restricted with this command, access to CatOS devices can only be achieved through a few trusted end-stations. This setup can be a hindrance in troubleshooting. Keep in mind that it is possible to spoof IP addresses and to fool filtered access, so this is only the first layer of protection.

Port Security

Consider utilizing port security in order to permit only one or several known MAC addresses to pass data on a particular port (in order to stop static end stations from being swapped for new stations without change control, for example). This is possible by with static MAC addresses.

```
set port security mod/port enable MAC address
```

This is also possible by learning restricted MAC addresses dynamically.

```
set port security port range enable
```

These options can be configured:

- **set port security *mod/port* age *time value*** —specifies the duration for which addresses on the port are secured before a new address can be learned. Valid time in minutes is 10 - 1440. Default is no aging.
- **set port security*mod/port* maximum *value*** —keyword that specifies the maximum number of MAC addresses to secure on the port. Valid values are 1 (default) - 1025.
- **set port security *mod/port* violation shutdown** —shuts down port (default) if violation occurs as well as sends syslog message (default) and discards the traffic.
- **set port security *mod/port* shutdown *time value*** —duration for which a port remains disabled. Valid values are 10 - 1440 minutes. Default is permanently shutdown

With CatOS 6.x and later, Cisco has introduced 802.1x authentication that allows clients to authenticate to a central server before ports can be enabled for data. This feature is in the early stages of support on such platforms as Windows XP, but can be considered a strategic direction by many enterprises. Refer to Configuring Port Security for information on how to configure port security on switches that run Cisco IOS Software.

Login Banners

Create appropriate device banners to state specifically the actions taken for unauthorized access. Do not advertise site name or network data that could provide information to unauthorized users. These banners provide recourse in case a device is compromised and the perpetrator is caught:.

```
# set banner motd ^C
*** Unauthorized Access Prohibited ***
***  All transactions are logged   ***
------------ Notice Board ------------
----Contact Joe Cisco at 1 800 go cisco for access problems----
^C
```

Physical Security

Devices must not be accessible physically without proper authorization, so the equipment must be in a controlled (locked) space. in order to ensure that the network stays operational and unaffected by malicious tampering of environmental factors, all equipment must have proper UPS (with redundant sources where possible) and temperature control (air conditioning). Remember, if physical access is breached by a person with malicious intent, disruption through password recovery or other methods is much more likely.

Terminal Access Controller Access Control System

By default, non-privileged and privileged mode passwords are global and apply to every user that accesses the switch or router, either from the console port or through a Telnet session across the network. Their implementation on network devices is time-consuming and non-centralized. It is also difficult to implement access restrictions using access lists that can be prone to configuration errors.

Three security systems are available to help control and police access to network devices. These use client/server architectures to place all security information in a single central database. These three security systems are:

- TACACS+
- RADIUS
- Kerberos

TACACS+ is a common deployment in Cisco networks and is the focus of this chapter. It provides these features:

- Authentication—the identification and verification process for a user. Several methods can be used to authenticate a user, but the most common includes a combination of user name and password.
- Authorization—of various commands can be granted once a user is authenticated.
- Accounting—the recording what a user is doing or has done on the device.

Refer to Configuring TACACS+, RADIUS, and Kerberos on Cisco Catalyst Switches for more details.

Operational Overview

The TACACS+ protocol forwards usernames and passwords to the centralized server, encrypted over the network using **MD5** one-way hashing (RFC 1321). It uses TCP port 49 as its transport protocol; this offers these advantages over UDP (used by RADIUS):

- Connection oriented transport

- Separate acknowledgement that a request has been received (TCP ACK), regardless of how loaded the backend authentication mechanism is currently
- Immediate indication of a server crash (RST packets)

During a session, if additional authorization checking is needed, the switch checks with TACACS+ to determine if the user is granted permission to use a particular command. This provides greater control over the commands that can be executed on the switch while de-coupling from the authentication mechanism. Using command accounting, it is possible to audit the commands a particular user has issued while attached to a particular network device.

When a user attempts a simple ASCII login by authenticating to a network device with TACACS+, this process typically occurs:

- When the connection is established, the switch contacts the TACACS+ daemon to obtain a username prompt, which is then displayed to the user. The user enters a username, and the switch contacts the TACACS+ daemon in order to obtain a password prompt. The switch displays the password prompt to the user, who then enters a password that is also sent to the TACACS+ daemon.
- The network device eventually receives one of these responses from the TACACS+ daemon:ACCEPT—the user is authenticated and service can begin. If the network device is configured to require authorization, authorization begins at this time.REJECT—the user has failed to authenticate. The user can be denied further access or is prompted to retry the login sequence depending on the TACACS+ daemon.ERROR—an error occurred at some time during authentication. This can be either at the daemon or in the network connection between the daemon and the switch. If an ERROR response is received, the network device typically tries to use an alternative method in order to authenticate the user.CONTINUE—the user is prompted for additional authentication information.
- Users must first successfully complete TACACS+ authentication before they proceed to TACACS+ authorization.
- If TACACS+ authorization is required, the TACACS+ daemon is again contacted and returns an ACCEPT or REJECT authorization response. If an ACCEPT response is returned, the response contains data in the form of attributes that are used to direct the EXEC or NETWORK session for that user, and determine the commands that the user can access.

Recommendation

Cisco recommends the use of TACACS+, as it can be easily implemented using CiscoSecure ACS for NT, Unix, or other third-party software. TACACS+ features include detailed accounting to provide statistics on command usage and system usage, MD5 encryption algorithm, and administrative control of authentication and authorization processes.

In this example, login and enable modes use the TACACS+ server for Authentication and can fall back to local authentication if the server is unavailable. This is an important back door to leave in most networks. Issue these commands in order to set up TACACS+:

```
set tacacs server server IP primary set tacacs server server IP
!--- Redundant servers are possible. set tacacs attempts 3
!--- This is the default. set tacacs key key
!--- MD5 encryption key. set tacacs timeout 15
!--- Longer server timeout (5 is default). set authentication login tacacs enable
```

```
set authentication enable tacacs enable
set authentication login local enable
set authentication enable local enable
!--- The last two commands are the default; they allow fallback !--- to local if no
TACACS+ server available.
```

Other Options

It is possible to use TACACS+ authorization to control the commands each user or user-group can execute on the switch, but it is difficult to make a recommendation because all customers have individual requirements in this area. Refer to Controlling Access to the Switch Using Authentication, Authorization, and Accounting for more information.

Finally, accounting commands provide an audit trail of what each user typed and configured. This is an example using the common practice of receiving the audit information at the end of the command:

```
set accounting connect enable start-stop tacacs+
set accounting exec enable start-stop tacacs+
set accounting system enable start-stop tacacs+
set accounting commands enable all start-stop tacacs+
set accounting update periodic 1
```

This configuration has these features:

- The **connect** command enables accounting of outbound connection events on the switch such as Telnet.
- The **exec** command enables accounting of login sessions on the switch such as operations staff.
- The **system** command enables accounting of system events on the switch such as reload or reset.
- The **commands** command enables accounting of what was entered on the switch, for both **show** and configuration commands.
- Periodic *updates* every minute to the server are helpful in order to record whether users are still logged in.

Configuration Checklist

This section provides a summary of the recommended configurations, excluding security details.

It is extremely helpful to label all ports.. Issue this command in order to label the ports:

```
set port description descriptive name
```

Use this key in conjunction with the Command tables listed:

Key:
Bold text - recommended change
Normal text - default, recommended setting

Global Configuration Commands

Command	Comment
set vtp domain *name* password*x*	Protect against unauthorized VTP updates from new switches.
set vtp mode transparent	Select VTP mode promoted in this document. Refer to the VLAN Trunking Protocol section of this document for more details.
set spantree enable all	Ensure that STP is enabled on all VLANs.
set spantree root *vlan*	Recommended to position root (and secondary root) bridges per VLAN.
set spantree backbonefast enable	Enable rapid STP convergence from indirect failures (only if all switches in domain support the feature).
set spantree uplinkfast enable	Enable rapid STP convergence from direct failures (for access layer switches only).
set spantree portfast bpdu-guard enable	Enable port to be shut down automatically if there is an unauthorized Spanning Tree extension.
set udld enable	Enable unidirectional link detection (need port level configuration as well).
set test diaglevel complete	Enable full diagnostics at boot up (default on Catalyst 4500/4000).
set test packetbuffer sun 3:30	Enable port buffer error checking (applies to Catalyst 5500/5000 only).
set logging buffer 500	Maintain maximum internal syslog buffer.
set logging server *IP address*	Configure target syslog sever for external system message logging.
set logging server enable	Allow the external logging server.
set logging timestamp enable	Enable timestamps of messages in the log.
set logging level spantree 6 default	Increase default STP syslog level.

set logging level sys 6 default	Increase default System syslog level.
set logging server severity 4	Allow the export of the higher severity syslog only.
set logging console disable	Disable the console unless troubleshooting.
set snmp community read-only *string*	Configure the password to allow remote data collection.
set snmp community read-write *string*	Configure the password to allow remote configuration.
set snmp community read-write-all *string*	Configure the password to allow remote configuration including passwords.
set snmp trap enable all	Enable SNMP traps to the NMS server for fault and event alerts.
set snmp trap *server address string*	Configure the address of the NMS trap receiver.
set snmp rmon enable	Enable RMON for local statistic gathering. Refer to the Remote Monitoring section of this document for more details.
set ntp broadcastclient enable	Enable accurate system clock reception from an upstream router.
set ntp timezone *zone name*	Set the local timezone for the device.
set ntp summertime *date change details*	Configure summertime if applicable for the timezone.
set ntp authentication enable	Configure encrypted time information for security purposes.
set ntp key *key*	Configure the encryption key.
set cdp enable	Ensure neighbor discovery is enabled (enabled on ports by default as well).
set tacacs server *IP address* primary	Configure the address of the AAA server.
set tacacs server *IP address*	Redundant AAA servers if possible.
set tacacs attempts 3	Allow 3 password attempts for the AAA user account.
set tacacs key *key*	Set the AAA MD5 encryption key.
set tacacs timeout 15	Allow longer server timeout (five seconds is default).
set authentication login tacacs enable	Use AAA for authentication for login.
set authentication enable tacacs enable	Use AAA for authentication for enable mode.

set authentication login local enable	Default; allows fallback to local if no AAA server available.
set authentication enable local enable	Default; allows fallback to local if no AAA server available.

Host Ports Configuration Commands

Command	Comment
set port host *port range*	Remove unnecessary port processing. This macro sets spantree PortFast enable, channel off, trunk off.
set udld disable *port range*	Remove unnecessary port processing (disabled on copper port by default).
set port speed *port range* auto	Use auto negotiation with up-to-date host NIC drivers.
set port trap *port range* disable	No need for SNMP traps for general users; track key ports only.

Server Configuration Commands

Command	Comment
set port host *port range*	Remove unnecessary port processing. This macro sets spantree PortFast enable, channel off, trunk off.
set udld disable *port range*	Remove unnecessary port processing (disabled on copper port by default).
set port speed *port range 10 | 100*	Usually configure static/server ports; otherwise, use autonegotiation.
set port duplex *port range full | half*	Usually static/server ports; otherwise, use autonegotiation.
set port trap *port range* enable	Key service ports must send trap to NMS.

Unused Ports Configuration Commands

Command	Comment
set spantree portfast *port range* disable	Enable necessary port processing and protection for STP.
set port disable *port range*	Disable unused ports.
set vlan *unused dummy vlan port*	Direct unauthorized traffic to unused VLAN if the port is

range	enabled.
set trunk *port range* off	Disable port from trunking until administered.
set port channel *port range* mode off	Disable port from channeling until administered.

Infrastructure Ports (switch-switch, switch-router)

Command	Comment
set udld enable *port range*	Enable unidirectional link detection (not default on copper ports).
set udld aggressive-mode enable *port range*	Enable aggressive mode (for devices that support it).
set port negotiation *port range*enable	Allow default GE autonegotiation of link parameters.
set port trap *port range* enable	Allow SNMP traps for these key ports.
set trunk *port range* off	Disable feature if not using trunks.
set trunk *mod/port desirable ISL* | *dot1q* | *negotiate*	If using trunks, dot1q is preferred.
clear trunk *mod/port vlan range*	Limit STP diameter by pruning VLANs from trunks where they are not needed.
set port channel *port range* mode off	Disable feature if not using channels.
set port channel *port range* mode desirable	If using channels, this enables PAgP.
set port channel all distribution ip both	Allow L3 source/destination load balancing if using channels (default on Catalyst 6500/6000).
set trunk *mod/port* nonegotiate *ISL* | *dot1q*	Disable DTP if trunking to router, Catalyst 2900XL, 3500, or other vendor.
set port negotiation *mod/port* disable	Negotiation can be incompatible for some old GE devices.

Related Information

- **Common CatOS Error Messages on Catalyst 4500/4000 Series Switches**
- **Common CatOS Error Messages on Catalyst 5000/5500 Series Switches**
- **Common CatOS Error Messages on Catalyst 6500/6000 Series Switches**
- **Switches Product Support**

- **LAN Switching Technology Support**
- **Technical Support & Documentation - Cisco Systems**