Cisco Nexus Validation Test (NVT)

March 13, 2013

Americas Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
http://www.cisco.com
Tel: 408 526-4000
800 553-NETS (6387)
Fax: 408 527-0883

Text Part Number: OL-27711-04

THE SPECIFICATIONS AND INFORMATION REGARDING THE PRODUCTS IN THIS MANUAL ARE SUBJECT TO CHANGE WITHOUT NOTICE. ALL STATEMENTS, INFORMATION, AND RECOMMENDATIONS IN THIS MANUAL ARE BELIEVED TO BE ACCURATE BUT ARE PRESENTED WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED. USERS MUST TAKE FULL RESPONSIBILITY FOR THEIR APPLICATION OF ANY PRODUCTS.

THE SOFTWARE LICENSE AND LIMITED WARRANTY FOR THE ACCOMPANYING PRODUCT ARE SET FORTH IN THE INFORMATION PACKET THAT SHIPPED WITH THE PRODUCT AND ARE INCORPORATED HEREIN BY THIS REFERENCE. IF YOU ARE UNABLE TO LOCATE THE SOFTWARE LICENSE OR LIMITED WARRANTY, CONTACT YOUR CISCO REPRESENTATIVE FOR A COPY.

The Cisco implementation of TCP header compression is an adaptation of a program developed by the University of California, Berkeley (UCB) as part of UCB’s public domain version of the UNIX operating system. All rights reserved. Copyright © 1981, Regents of the University of California.

NOTWITHSTANDING ANY OTHER WARRANTY HEREIN, ALL DOCUMENT FILES AND SOFTWARE OF THESE SUPPLIERS ARE PROVIDED “AS IS” WITH ALL FAULTS. CISCO AND THE ABOVE-NAMED SUPPLIERS DISCLAIM ALL WARRANTIES, EXPRESSED OR IMPLIED, INCLUDING, WITHOUT LIMITATION, THOSE OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT OR ARISING FROM A COURSE OF DEALING, USAGE, OR TRADE PRACTICE.

IN NO EVENT SHALL CISCO OR ITS SUPPLIERS BE LIABLE FOR ANY INDIRECT, SPECIAL, CONSEQUENTIAL, OR INCIDENTAL DAMAGES, INCLUDING, WITHOUT LIMITATION, LOST PROFITS OR LOSS OR DAMAGE TO DATA ARISING OUT OF THE USE OR INABILITY TO USE THIS MANUAL, EVEN IF CISCO OR ITS SUPPLIERS HAVE BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

Cisco and the Cisco Logo are trademarks of Cisco Systems, Inc. and/or its affiliates in the U.S. and other countries. A listing of Cisco's trademarks can be found at www.cisco.com/go/trademarks. Third party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (1005R)

Any Internet Protocol (IP) addresses and phone numbers used in this document are not intended to be actual addresses and phone numbers. Any examples, command display output, network topology diagrams, and other figures included in the document are shown for illustrative purposes only. Any use of actual IP addresses or phone numbers in illustrative content is unintentional and coincidental.

Cisco Nexus Validation Test (NVT)

CONTENTS

Preface v

Audience v

Change History v

Organization v

Related Documentation vi

Document Conventions vi

CHAPTER 1 Introduction 1-1

Purpose and Scope of Nexus Validation Test 1-1

Platforms and Releases 1-1

Description of the Test Network 1-2

Hierarchical Network Model 1-2

Topology of the Test Network 1-3

Edge/Core Layer 1-3

Aggregation Layer 1-4

Access Layer 1-4

Test Network Configuration 1-4

Features Deployed and Verified 1-5

Edge Layer 1-5

Core Layer 1-6

Aggregation Layer 1-7

CHAPTER 2 Methodology 2-1

Test Cycle 2-1

Network Disruption Test Cases 2-1

System Level 2-2

Edge Layer 2-2

Core Layer 2-4

Aggregation Layer 2-4

Access Layer 2-6

Sample Test Case 2-7

CHAPTER 3 NVT Summary, Issues, and Recommendations 3-1

Summary 3-1

Issues and Recommendations 3-1

Optimization of CoPP 3-2

BFD or Aggressive Timers 3-3

ISSU 3-3

Routing Protocols and MTU 3-4

Reserved VLANs 3-4

Link Aggregation Protocols 3-5

Multichassis EtherChannel Interoperability 3-5

IS-IS Default Metric Style 3-5

OSPF Interface Metric 3-5

BGP Per-Interface Fast External Failover 3-5

Multicast Multipath 3-5

Multicast SPT Threshold 3-6

Dynamic Trunking Protocol 3-6

Conclusion 3-6

GLOSSARY

Preface

This preface describes the audience and organization of the *Cisco Nexus Validation Test (NVT)*, and provides document conventions and information on how to obtain related documentation.

NVT examines features and configuration parameters commonly used in networks deployed with the Cisco Catalyst 6500 Series Switch. These features are then validated on the Cisco Nexus 7000 Series Switch deployed in similar networks to achieve the same capabilities and services. Any additional configuration or differences encountered during testing are noted in this document.

This preface contains the following topics:

- Audience, page v
- Change History, page v
- Organization, page v
- Related Documentation, page vi
- Document Conventions, page vi

Audience

This document is intended for network planners, engineers, and managers who are deploying the Cisco Nexus 7000 Series Switch in a datacenter as a companion to or replacement for the Cisco Catalyst 6500 Series Switch.

Change History

Date	Description
November 9, 2012	
March 13, 2013	Added "Conclusion" section on page 6 in Chapter 3, "NVT Summary, Issues, and Recommendations"

Organization

This document is organized into the following chapters:

- Chapter 1, “Introduction”
- Chapter 2, “Methodology”
- Chapter 3, “NVT Summary, Issues, and Recommendations”

Related Documentation

Related documentation for the Cisco Nexus 7000 Series Switch and Cisco IOS is available at the following links:

- *Cisco Nexus 7000 Series Switches Documentation Roadmap*
- *Cisco Nexus 7000 NX-OS/IOS Comparison Tech Notes*
- *Release Notes for the Cisco Nexus 7000 Series Switch*
- *Release Note for the Catalyst 4500 Series Switch, Cisco IOS Releases 12.2(54)SG to 12.2(40)SG*
- *Catalyst 4500 Series Switch Software Configuration Guide, 12.2(46)SG*
- *Release Notes for the Catalyst 4500E Series Switch, Cisco IOS XE 3.3.0SG*
- *Catalyst 4500 Series Switch Software Configuration Guide, Release IOS XE 3.3.0SG and IOS 15.1(1)SG*
- *Configuration Guides, Cisco IOS 15.0SY*
- *Release Notes for Cisco IOS Release 12.2SX*
- *Configuration Guides, Cisco IOS 12.2SX*

Document Conventions

Command descriptions use these conventions:

Convention	Description
boldface font	Commands and keywords are in boldface.
italic font	Arguments for which you supply values are in italics.
[ ]	Elements in square brackets are optional.
[x I y I z]	Optional alternative keywords are grouped in brackets and separated by vertical bars.
string	A nonquoted set of characters. Do not use quotation marks around the string or the string will include the quotation marks.

Screen examples use these conventions:

`screen font`	Terminal sessions and information that the switch displays are in screen font.
`boldface screen font`	Information that you must enter is in boldface screen font.
`italic screen font`	Arguments for which you supply values are in italic screen font.
< >	Nonprinting characters, such as passwords, are in angle brackets.

[ ]	Default responses to system prompts are in square brackets.
!, #	An exclamation point (!) or a pound sign (#) at the beginning of a line of code indicates a comment line.

This document uses the following conventions:

Note Means reader *take note*. Notes contain helpful suggestions or references to material not covered in the manual.

Caution Means *reader be careful*. In this situation, you might do something that could result in equipment damage or loss of data.

CHAPTER 1

Introduction

This chapter provides an overview of the purpose and methodology of this document. This chapter contains the following sections:

- Purpose and Scope of Nexus Validation Test, page 1-1
- Platforms and Releases, page 1-1
- Description of the Test Network, page 1-2
- Features Deployed and Verified, page 1-5

Purpose and Scope of Nexus Validation Test

Cisco Nexus 7000 hardware and software releases must pass Cisco's comprehensive quality assurance process, which includes a multistage approach comprising extensive unit test, feature test, and system-level test. Each successive stage in the process adds increasingly higher levels of complexity in a multidimensional mix of features and topology.

NVT has been established as an additional quality assurance stage in order to leverage customer feedback and requirements into the product development cycle. NVT will validate and publish guidelines for deploying Nexus 7000 and NX-OS solutions for datacenter networks.

This document describes the first phase of NVT, which includes IOS to NX-OS migration. The Nexus 7000 is featured throughout the multi-tiered test network topology at every Places-in-the-Network (PIN). IOS on Catalyst 6k and 4k are also deployed side-by-side with the Nexus 7000 to ensure that customers with IOS experience will be able to successfully transition to NX-OS platforms. NVT will provide recommendations and guidelines to ensure that the hardware and features used in this network will behave and perform similarly between the platforms.

Platforms and Releases

The migration testing described in this document incorporates the following platforms and releases:

Platform	Software Release
Nexus 7000	NX-OS 5.2(5)
Catalyst 6500 Supervisor 720 and 720-10G	IOS 12.2(33)SXJ
Catalyst 6500 Supervisor 2T	IOS 15.0(1)SY

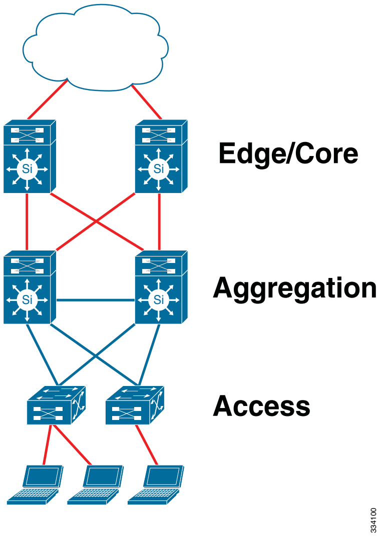
Platform	Software Release
Catalyst 4500 Supervisor 7E	IOS-XE 3.3.0SG
Catalyst 4948	IOS 12.2(46)SG

Note Any reference to IOS or NX-OS in this document applies only to the software versions and platforms mentioned in the preceding table.

Description of the Test Network

Hierarchical Network Model

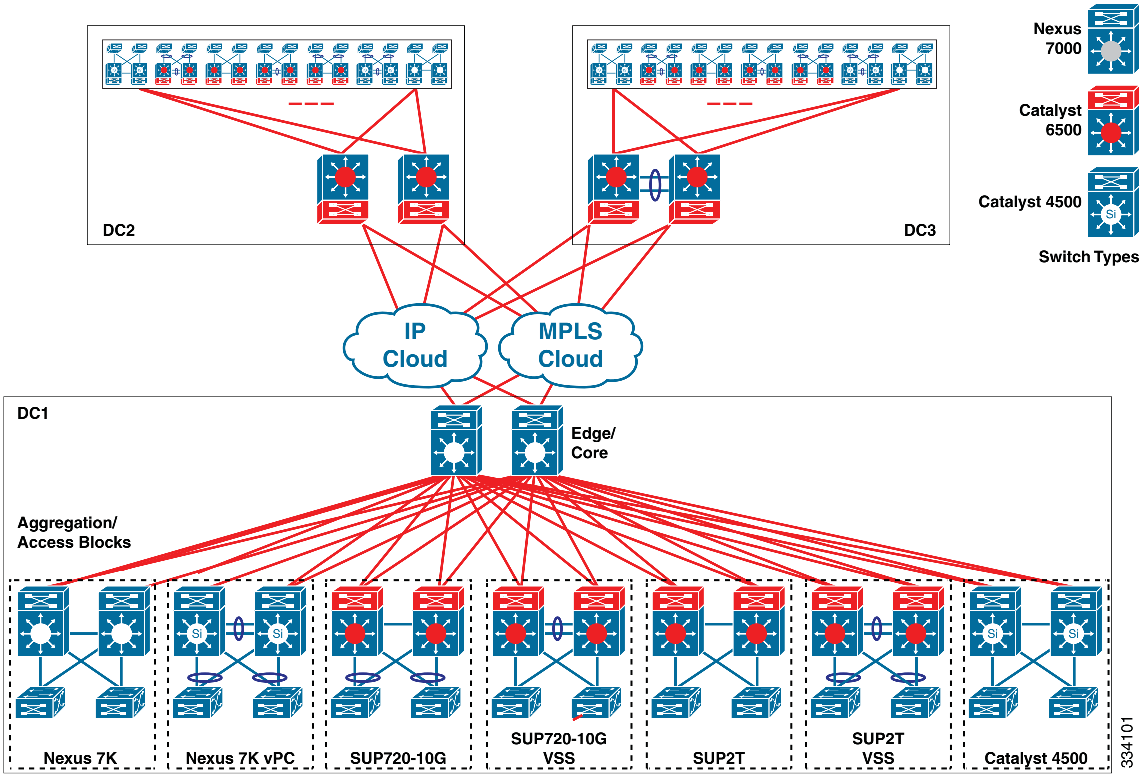
The test network is a traditional hierarchical model comprising the edge, core, aggregation, and access layers, as shown in the following figure.



Because most features of interest in the initial phase of testing involve the aggregation layer, the edge and core layers are collapsed into a single layer.

Topology of the Test Network

The following figure illustrates the test network topology, consisting of three datacenter sites interconnected through a public IP cloud and an MPLS cloud. The mixture of cross-platform devices presented in this topology is intended for the purpose of migration and interoperability testing.



Edge/Core Layer

The edge layer provides connectivity and security between the datacenter and the public network. This layer also provides private extensions between datacenters through the public cloud. In the test network, private extensions between the datacenters are implemented with GRE tunnels and MPLS VPNs at the edge.

The edge/core layer provides routing and high bandwidth connectivity between the edge and the aggregation layers.

The edge layer of each datacenter in the test network is implemented using one of the following three platform types to ensure feature parity and interoperability:

- Cisco Nexus 7000 Series Switch
- Cisco Catalyst 6500 Series Switch
- Cisco Catalyst 6500 Series Switch Virtual Switching System (VSS)

The use of three different platforms at the edge/core allows for the comparison of feature behavior, performance, and scale between Catalyst 6500 and Nexus 7000 systems operating at the edge/core layers.

Aggregation Layer

The aggregation layer provides connectivity and policy services for traffic flows of all switches within the access-aggregation block.

The aggregation layer of each datacenter consists of seven blocks, implemented using each of these seven platforms:

- Block 1: Cisco Nexus 7000 Series Switch
- Block 2: Cisco Nexus 7000 Series Switch with virtual port channel (vPC)
- Block 3: Cisco Catalyst 6500 Series Switch Supervisor Engine 720
- Block 4: Cisco Catalyst 6500 Series Switch Supervisor Engine 720-10G VSS
- Block 5: Cisco Catalyst 6500 Series Switch Supervisor Engine 2T
- Block 6: Cisco Catalyst 6500 Series Switch Supervisor Engine 2T VSS
- Block 7: Cisco Catalyst 4500 Series Switch

The aggregation layer of each datacenter is identical in design to the others to ensure that each of the core platforms interoperates well with all major Cisco modular switching products. The common design allows for the comparison of feature behavior, performance, and scale among the major Cisco modular switching products operating at the aggregation layers.

Access Layer

The access layer provides connectivity to a scaled number of end devices in the datacenter network.

In the test network, the access layer is deployed using various Cisco Catalyst switching products.

The migration of the access layer from IOS to NX-OS devices is not covered in this phase of NVT.

Test Network Configuration

The following configuration details are applied to the test network:

- In order to maximize the number of features and protocols that can be tested in parallel, MPLS Multi-VRF (VRF-lite) is deployed across all datacenters.
- Each datacenter has four VPN routing and forwarding instances (VRFs), with each VRF running one of the following protocols for unicast routing: OSPF, EIGRP, IS-IS, and BGP. For multicast routing, each VRF runs PIM ASM, Bidir, and SSM.

Note For vPC configuration, PIM Bidir and SSM are not supported and are therefore not tested.

- The Nexus 7000 are further virtualized at the device level using Virtual Device Contexts (VDC).
- The entire test network is configured to support SSO/NSF.
- The test network is configured and operating in both IPv4 and IPv6 modes. In this phase, the primary focus of the test cases is IPv4.

- Bidirectional Forwarding Detection (BFD) is recommended on the Nexus 7000 to optimize network peer failure detection. However, within the test network topology, aggressive timers for routing protocols are used on interfaces where BFD is not supported. For example, BFD is not supported on port-channels and SVIs on Catalyst switches. Otherwise, BFD is used with protocol clients running with default timers.

  PIM and First Hop Redundancy Protocol (FHRP) are not supported as BFD clients on the Catalyst 6500 and 4500. Within the NVT topology, even though BFD is configured on routed interfaces between Catalyst switches and the Nexus 7000, the multicast routing protocols were not tested as BFD clients. Therefore, PIM is tested with aggressive timers on those interfaces.

  The BFD retransmit interval is configured to be 1 second with 3x holddown multiplier. These parameters are chosen to match the protocols running aggressive timers with 1 second hello intervals.

Note Disable Internet Control Message Protocol (ICMP) redirect messages on BFD-enabled interfaces.

Features Deployed and Verified

Features relevant to the datacenter edge, core, aggregation, and access layers were tested and verified. Any issues affecting operation and differences between NX-OS and IOS that were not resolved are noted in the subsequent sections.

Edge Layer

Route-Map Based Policies

From the data center edge to the IP and MPLS clouds, route-map based policies control route redistribution between the datacenter IGP and BGP peerings.

The configuration of BGP neighbors is grouped and ordered differently between IOS and NX-OS. Similarly, configurations for other features are more hierarchically structured in NX-OS relative to IOS.

Note In the route-map definition, NX-OS supports only prefix lists, while IOS supports prefix lists and access lists.

Security ACLs

Security ACLs are used to regulate data flows and control-plane traffic entering and leaving the data centers.

The storage of ACE sequence numbers differs between IOS and NX-OS:

- IOS—The access-list definition does not retain and store the ACE sequence numbers in persistent storage; however, the configured sequence is maintained.
- NX-OS—The ACE sequence numbers are stored as a part of the startup configuration.

The following example shows that the IOS system (Cat-6500) does not store the ACE sequence numbers in the configuration while the NX-OS system (Nexus-7000) stores the numbers.

```
Cat-6500# show ip access-lists nvt
```

```
Extended IP access list nvt
    10 permit ip any 230.81.1.0 0.0.0.255
    20 permit ip any 230.82.1.0 0.0.0.255
Cat-6500# show running-config | begin nvt
ip access-list extended nvt
permit ip any 230.81.1.0 0.0.0.255      <- Sequence number is not retained
permit ip any 230.82.1.0 0.0.0.255


Nexus-7000# show ip access-lists nvt
IP access list nvt
        10 permit ip any 230.81.1.0/24
        20 permit ip any 230.82.1.0/24
Nexus-7000# show running-config | begin nvt
ip access-list nvt
  10 permit ip any 230.81.1.0/24      <- Sequence number is retained
  20 permit ip any 230.82.1.0/24
```

NetFlow Data Export

Support for NetFlow data export versions differs between the tested switching platforms:

- Catalyst Supervisor Engine 720 supports NetFlow data export versions 5 and 7.
- Catalyst Supervisor Engine 2T supports NetFlow data export versions 5 and 9.
- Catalyst 4500 supports NetFlow data export versions 5 and 9.
- Nexus 7000 supports NetFlow data export versions 5 and 9.

NetFlow data export version 9 allows for Flexible NetFlow.

GRE and MLPS VPNs

The data centers are interconnected over the public clouds using GRE and MPLS VPNs. For the Nexus 7000 edge, multicast traffic between the data centers is carried only over GRE.

- IOS—The MTU of the GRE tunnel interface is automatically derived from the transport interface; there is no option to configure the tunnel MTU.
- NX-OS—The MTU should be manually configured to match the value of the tunnel destination.

Core Layer

Scalability

The core layer is configured to support four VRFs. Each VRF learns up to 5000 unicast routes from the edge peers connected to the public cloud. A small subset of those unicast routes are distributed into each of the aggregation blocks. Although the test topology contains only seven aggregation blocks, up to 13 additional aggregation blocks were simulated for unicast routing scale.

Each of the four core layer VRFs also learns up to 2000 multicast routes from the seven aggregation blocks and the other two datacenters through the GRE tunnels.

PIM Rendezvous Point and MSDP

For Any Source Multicast (ASM), the core layer serves as Multicast Source Discovery Protocol (MSDP) Anycast RP. The sa-cache table was tested up to 7500 entries. For multicast Bidir, the core layer serves as Phantom RP.

The MSDP default configuration differs between IOS and NX-OS:

- IOS—MSDP sa-cache must be explicitly configured using the **ip msdp cache-sa-state** command.
- NX-OS—MSDP sa-cache is enabled by default.

Aggregation Layer

Layer 2 Forwarding

The aggregation layer provides loop-free layer 2 access to end devices. The Nexus 7000 aggregation blocks are built with STP and vPC, while the Catalyst 6500 blocks use STP and VSS. The vPC and VSS topologies cover orphan and non-orphan scenarios.

The default MAC address aging time differs between IOS and NX-OS:

Platform	Default MAC Address Aging Time, in seconds
IOS 12.2(33)SXJ	300
IOS 15.0(1)SY	480
NX-OS	1800

FHRP and ARP/ND

The aggregation layer participates in unicast and multicast routing with the core layer for all VRFs. For unicast routing, this layer provides FHRP for gateway services to end devices with ARP/ND operations.

- IOS—In VSS systems there is only one control-plane and all the forwarding engines in both chassis are programmed by this single control plane; for this reason, FHRP is not strictly required.
- NX-OS—With vPC, the hardware forwarding engines on both vPC peers are programmed to be Active/Active even though the control plane will stay in Active/Standby mode.

The default ARP timeout differs between IOS and NX-OS:

Platform	Default ARP Timeout, in seconds
IOS	14400
NX-OS	1500

Note Cisco recommends that you configure the ARP timeout to be slightly shorter than the MAC address aging time to minimize flooding due to host inactivity.

IGMP/MLD

The aggregation layer provides IGMP/MLD Snooping and Querier with last hop routing. This layer is also the first hop router for multicast data sources and provides PIM ASM source registration. For a limited set of multicast groups with directly connected sources, this layer provides MSDP Anycast RP services.

The default IGMP Querier interval differs between IOS and NX-OS:

Platform	Default IGMP Querier Interval, in seconds
IOS	60
NX-OS	125

PIM ASM

- For Any Source Multicast (ASM), each aggregation block is configured with the SPT threshold set to “infinity” except for the vPC block, where the setting is not supported.
- PIM rendezvous points (RPs) are located at the core layer and at each aggregation block.
- The groups registered to the RP located at each aggregation block are originated from sources attached to the access switches within that aggregation block.
- The groups registered to the RP located at the core layer are originated from sources attached to every aggregation block.
- Multicast receivers are located at each aggregation block and these receivers join to all multicast groups.
- Multicast multipath routing is enabled across the entire network on the Catalyst switches to match the default behavior on the Nexus 7000.

CHAPTER 2

Methodology

This chapter contains the following sections:

- Test Cycle, page 2-1
- Sample Test Case, page 2-7

Test Cycle

The test cycle consists of the following steps:

1. Network configuration and verification
2. Image upgrade and rollback with ISSU
3. Induce network disruptions at each layer (see Network Disruption Test Cases, page 2-1)
4. Scale control-plane peers and routes for both unicast and multicast routing at the core layer
5. Scale the number of access level switches and hosts to stress ARP and IGMP at the aggregation layer
6. Extended uptime monitoring to check for CPU and memory usage anomalies

Network Disruption Test Cases

The following sections describe the test disruptions and the verification criteria:

- System Level, page 2-2
- Edge Layer, page 2-2
- Core Layer, page 2-4
- Aggregation Layer, page 2-4
- Access Layer, page 2-6

System Level

Disruption	Verification
Image upgrade and rollback with ISSU	Hitless upgrade/rollback for all configured features with parallel enhancement

Edge Layer

Disruption	Verification
Router Link Failure/Recovery between Edge and Public Cloud	• BGP reconvergence (control-plane & data-plane) • IGP and Multicast services reconvergence (control-plane & data plane) • MPLS/VPN and LDP reconvergence (control-plane & data-plane) • BFD peer detection and client notifications • GRE Tunnel re-route due to transport disruption
Member of Port-channel Failure/Recovery between Edge and Public Cloud	• Traffic load-sharing for port-channels • LACP interoperability • Unidirectional Link Detection (UDLD)
Clear IGP Neighbors/Process at Edge	Stress test for control-plane recovery
Clear IPv4/IPv6 Unicast Routes at Edge	Stress test for control-plane recovery
Clear IPv4/IPv6 Multicast Routes at Edge	Stress test for control-plane recovery
Edge Switch System Failure/Recovery	• BGP reconvergence (control-plane & data-plane) • IGP and Multicast services reconvergence (control-plane & data plane) • MPLS/VPN and LDP reconvergence (control-plane & data-plane) • BFD peer detection and client notifications • GRE Tunnel re-route due to transport disruption • VDC failure does not impact other VDCs
Edge Switch Power Redundancy	Partial Power loss causes no impact to control/data plane
Edge Switch Supervisor High-Availability	• SSO/NSF, in-chassis and on peers • SSO/NSF interoperability

Disruption	Verification
Edge Switch Fabric High-Availability	Fabric module failure causes no impact to control/data plane
Line Card OIR at Edge Switch	• Hitless operation for non-affected ports • Traffic load-sharing for distributed port-channels • IGP and PIM reconvergence (control-plane & data plane) • BFD peer detection and client notifications • LACP interoperability for distributed port-channels • Unidirectional Link Detection (UDLD)

Core Layer

Disruption	Verification
Router Link Failure/Recovery between Core and Edge	• IGP and PIM reconvergence (control-plane & data plane) • BFD peer detection and client notifications
Member of Port-channel Failure/Recovery between Core and Edge	• Traffic load-sharing for port-channels • LACP interoperability • Unidirectional Link Detection (UDLD)
Clear IGP Neighbors/Process at Core	Stress test for control-plane recovery
Clear IPv4/IPv6 Unicast Routes at Core	Stress test for control-plane recovery
Clear IPv4/IPv6 Multicast Routes at Core	Stress test for control-plane recovery
Core Switch System Failure/Recovery	• IGP and PIM reconvergence (control-plane & data plane) • BFD peer detection and client notifications • PIM Rendezvous Point redundancy & Back-up verification • VDC failure does not impact other VDCs
Core Switch Power Redundancy	Partial Power loss causes no impact to control/data plane
Core Switch Supervisor High-Availability	• SSO/NSF, in-chassis and on peers • SSO/NSF interoperability
Core Switch Fabric High-Availability	Fabric module failure causes no impact to control/data plane
Line Card OIR at Core Switch	• Hitless operation for non-affected ports • Traffic load-sharing for distributed port-channels • IGP and PIM reconvergence (control-plane & data plane) • BFD peer detection and client notifications • LACP interoperability for distributed port-channels • Unidirectional Link Detection (UDLD)

Aggregation Layer

Disruption	Verification
Router Link Failure/Recovery between Aggregation and Core	• IGP and PIM reconvergence (control-plane & data plane) • BFD peer detection and client notifications
Member of Port-channel Failure/Recovery between Aggregation and Core	• Traffic load-sharing for port-channels • LACP interoperability • Unidirectional Link Detection (UDLD)

Disruption	Verification
Router Link Failure/Recovery between Aggregation and Access	• STP reconvergence • IGMP reprogramming with snooping • MAC address re-learning • Security ACL & FNF reprogramming • No FHRP impact • No ARP/ND impact • No BFD impact • vPC functionality
Member of Port-channel Failure/Recovery between Aggregation and Access	• Traffic load-sharing for port-channels • LACP interoperability • Unidirectional Link Detection (UDLD)
Clear IGP Neighbors/Process at Aggregation	Stress test for control-plane recovery
Clear IPv4/IPv6 Unicast Routes at Aggregation	Stress test for control-plane recovery
Clear IPv4/IPv6 Multicast Routes at Aggregation	Stress test for control-plane recovery
Aggregation Switch System Failure/Recovery	• STP reconvergence • IGP and PIM reconvergence (control-plane & data plane) • BFD peer detection and client notifications • PIM Rendezvous Point redundancy & Back-up verification • PIM DR/BDR functionality • IGMP Snooping & Querier functionality • VDC failure does not impact other VDCs • Security ACL & FNF reprogramming • FHRP redundancy • MAC address learning • ARP/ND re-learning • vPC functionality
Aggregation Switch Power Redundancy	Partial Power loss causes no impact to control/data plane
Aggregation Switch Supervisor High-Availability	• SSO/NSF, in-chassis and on peers • SSO/NSF interoperability • No impact to vPC peering status
Aggregation Switch Fabric High-Availability	Fabric module failure causes no impact to control/data plane

Disruption	Verification
Line Card OIR at Aggregation Switch	• Hitless operation for non-affected ports • Traffic load-sharing for distributed port-channels • IGP and PIM reconvergence (control-plane & data plane) • BFD peer detection and client notifications • LACP interoperability for distributed port-channels • Unidirectional Link Detection (UDLD)
vPC peer-link/keep-alive Failure/Recovery	vPC functionality and peering status
vPC Leg Failure/Recovery	• No impact to STP overlay • IGMP reprogramming with snooping • MAC address re-learning • Security ACL & FNF reprogramming • No FHRP impact • No ARP/ND impact
vPC Leg member Failure/Recovery	• Traffic load-sharing for port-channels • LACP interoperability • Unidirectional Link Detection (UDLD)

Access Layer

Disruption	Verification
Access/ToR Switch System Failure/Recovery	• STP reconvergence • IGMP snooping reprogramming • MAC address re-learning • No impact to other vPCs

Sample Test Case

Sample Test Case	
Title	Link failure between aggregation and core layers
Description	Verify network control and data plane recovery after link flap
Test Setup	• Reference topology • Reference network configuration setup test case • Reference test plan for control and data plane setup matrices
Procedure	**1.** Fail one of the links between the aggregation and core layers. **2.** Recover the above link. **3.** Repeat the same test at least 5 iterations to ensure consistent behavior for the devices and network. **4.** Repeat the above procedures for the other links between the aggregation and core layers.
Pass/Fail Criteria	• During the link failure, traffic should drop in proportion to the number of links and paths affected, and the traffic should be able to reconverge within the expected time relative to a previously-established Catalyst 6500 baseline. • Ensure that the unicast and multicast routing protocols have detected peer failure in order to start network reconvergence within the expected time. • Verify the convergence pattern is as expected. • Verify the CPU usage pattern is as expected. • Verify the memory usage is as expected. • Verify the route tables for both unicast and multicast routing are updated correctly on all switches in the network. Ensure that only affected switches show change in the forwarding tables. • Verify the hardware forwarding entries, line card programming, fabric programming, outgoing interface, forwarding engine entries, for both unicast and multicast routing are updated correctly on all switches in the network. • Verify Layer 2 forwarding tables on aggregation and access switches. They should not be affected by this failure.

CHAPTER 3

NVT Summary, Issues, and Recommendations

This chapter contains the following sections:

- Summary, page 3-1
- Issues and Recommendations, page 3-1
- Conclusion, page 3-6

Summary

The test results verify that the services and features configured and tested (as described in this document) at each network layer of a running Cisco Catalyst platform can be deployed using similar feature sets on the Cisco Nexus 7000. In general, the results show that after any of the disruptions and subsequent recoveries, the network reconverges to the expected state within the expected time frame.

Issues and Recommendations

Specific issues and recommendations derived from the testing experience are discussed in this section. This section contains the following topics:

- Optimization of CoPP, page 3-2
- BFD or Aggressive Timers, page 3-3
- ISSU, page 3-3
- Routing Protocols and MTU, page 3-4
- Reserved VLANs, page 3-4
- Link Aggregation Protocols, page 3-5
- Multichassis EtherChannel Interoperability, page 3-5
- IS-IS Default Metric Style, page 3-5
- OSPF Interface Metric, page 3-5
- BGP Per-Interface Fast External Failover, page 3-5
- Multicast Multipath, page 3-5
- Multicast SPT Threshold, page 3-6
- Dynamic Trunking Protocol, page 3-6

Optimization of CoPP

CoPP optimization is recommended on the Nexus 7000 to achieve expected multicast routing scale and performance on the first hop source router for ASM. By default, any directly-connected multicast source will have its data traffic rate-limited by the Layer 3 multicast directly-connected rate limiter and further policed by the CoPP class default. This double layer of limiting may affect PIM source registration performance, especially when a large number of sources come online at the same time.

For NVT, the Layer 3 multicast directly-connected rate limiter is disabled. A new CoPP class is created to police multicast source data traffic so that multicast source registration performance with the NVT test profile is comparable between Nexus 7000 and Catalyst 6500 Supervisor Engine 2T. The ACL used in this CoPP class ensures that control plane protocol packets are not policed.

In order to update the CoPP configuration on the Nexus 7000, enter the following command to create a copy of the default configuration:

```
copp copy profile lenient prefix test
```

Enter the following commands to apply the copy to the control plane interface:

```
control-plane
  service-policy input test-copp-policy-lenient

hardware rate-limiter layer-3 multicast directly-connected disable
ip access-list multicast-source-data
  10 deny ip any 224.0.0.0/24
  20 deny ip any 224.0.1.0/24
  30 permit ip any 224.0.0.0/4

class-map type control-plane match-any multicast-source-data
  match access-group name multicast-source-data

policy-map type control-plane test-copp-policy-lenient
  class test-copp-class-critical
    set cos 7
    police cir 39600 kbps bc 375 ms conform transmit violate drop
  class test-copp-class-important
    set cos 6
    police cir 1060 kbps bc 1500 ms conform transmit violate drop
  class test-copp-class-management
    set cos 2
    police cir 10000 kbps bc 375 ms conform transmit violate drop
  class test-copp-class-normal
    set cos 1
    police cir 680 kbps bc 375 ms conform transmit violate drop
  class test-copp-class-normal-dhcp
    set cos 1
    police cir 680 kbps bc 375 ms conform transmit violate drop
  class test-copp-class-normal-dhcp-relay-response
    set cos 1
    police cir 900 kbps bc 750 ms conform transmit violate drop
  class test-copp-class-redirect
    set cos 1
    police cir 280 kbps bc 375 ms conform transmit violate drop
  class test-copp-class-exception
    set cos 1
    police cir 360 kbps bc 375 ms conform transmit violate drop
  class test-copp-class-monitoring
    set cos 1
    police cir 130 kbps bc 1500 ms conform transmit violate drop
  class test-copp-class-l2-unpoliced
    police cir 8 gbps bc 5 mbytes conform transmit violate transmit
```

```
class test-copp-class-undesirable
  set cos 0
  police cir 32 kbps bc 375 ms conform drop violate drop
class test-copp-class-l2-default
  police cir 100 kbps bc 375 ms conform transmit violate drop
class multicast-source-data
  police cir 1000 kbps bc 250 ms conform transmit violate drop
class class-default
  set cos 0
  police cir 100 kbps bc 250 ms conform transmit violate drop
```

Note On the Catalyst 6500 Supervisor Engine 720, the following rate limiters must be configured. In the absence of rate limiters, control plane protocols like BFD with short keepalive intervals or aggressive timers may flap.

```
mls rate-limit multicast ipv4 fib-miss 1000 100
mls rate-limit multicast ipv4 non-rpf 1000 100
mls rate-limit multicast ipv4 connected 1000 100
mls rate-limit multicast ipv4 partial 1000 100
mls rate-limit multicast ipv6 connected 1000 100
mls rate-limit multicast ipv6 mld 10 100
```

On the Catalyst 6500 Supervisor Engine 2T, there is no need for additional configuration to protect the control plane.

BFD or Aggressive Timers

BFD is recommended due to lower control plane overhead in order to achieve fast network failure detection and reconvergence. With BFD, any number of supported clients can piggy-back on top of one BFD session per connection for fast failure detection. The Nexus 7000 further enhances this capability with distributed BFD where BFD runs per line card instead of on the supervisor. The list of clients supporting BFD on the Nexus 7000 is extensive; however, the Catalyst 6500 and 4500 do not support BFD on all types of interfaces, and the list of BFD clients supported is less extensive. Some common examples of unsupported interfaces are port-channels, SVIs, and sub-interfaces; some examples of unsupported clients are the FHRP protocols and PIM. For connections between the Nexus 7000 and these platforms requiring fast peer failure detection for unsupported interfaces and unsupported BFD clients, aggressive timers are used.

The BFD retransmit interval is configured to be 1 second with 3x holddown multiplier. These parameters are chosen to match the protocols running aggressive timers with 1 second hello intervals.

ISSU

ISSU and rollback was performed between 5.2.5 and the following images:

- 5.2.4
- 5.2.3a
- 5.1.6
- 6.1.1

Upgrades and the rollbacks were tested with the ‘parallel’ option (where applicable) to minimize network maintenance window.

For networks running EIGRP, ISSU may cause routing peers to flap, especially for very high-scaled networks.

For switches running VTP in server mode, ISSU is recommended. Otherwise, an image version change via reload may cause VTP VLAN configuration to be lost.

Routing Protocols and MTU

If jumbo MTU is configured on the Nexus 7000, unicast routing protocols will leverage jumbo MTU for control plane update packets. PIM does not use jumbo MTU.

IOS on the Catalyst 6500 does not use jumbo MTU to send control plane protocol updates.

This is for information only. No additional configuration is required for the systems to fully interoperate.

Reserved VLANs

On the Catalyst 6500 Supervisor Engine 2T and the Nexus 7000, VLANs are not reserved for Layer 3 LAN ports and subinterfaces; however, some software features use internal VLANs in the extended range.

On the Catalyst 6500 Supervisor Engine 720 and Catalyst 4500/4948, VLANs are reserved for Layer 3 LAN ports and subinterfaces, and some software features use internal VLANs in the extended range.

You cannot use any VLAN that has been allocated for internal use.

For additional details, see the following references:

NX-OS 5.x releases

http://www.cisco.com/en/US/docs/switches/datacenter/sw/5_x/nx-os/layer2/configuration/guide/Cisco_Nexus_7000_Series_NX-OS_Layer_2_Switching_Configuration_Guide_Release_5.x_chapter4.html#task_67E5266F50104AF38E5149C1CC56B1A7

Catalyst 6500 SX releases

http://www.cisco.com/en/US/docs/switches/lan/catalyst6500/ios/12.2SX/configuration/guide/vlans.html#wpmkr1037585

Catalyst 6500 Sup2T 15.0SY releases

http://www.cisco.com/en/US/docs/switches/lan/catalyst6500/ios/15.0SY/configuration/guide/vlans.html#wpmkr1037585

Catalyst 4500 XE 3.3.0

http://www.cisco.com/en/US/docs/switches/lan/catalyst4500/15.1/XE_330SG/configuration/guide/vlans.html

Catalyst 4900 12.2(46)SG

http://www.cisco.com/en/US/docs/switches/lan/catalyst4500/12.2/46sg/configuration/guide/vlans.html

Link Aggregation Protocols

IOS—PAgP (auto/desirable) is supported.

NX-OS— PAgP (auto/desirable) is not supported.

Note NX-OS does not support PAgP. When a Catalyst 6500 VSS is connected to a Nexus 7000, enhanced PAgP for VSS Dual-Active Detection cannot be used on that link.

Multichassis EtherChannel Interoperability

To reduce data loss following a stateful switchover (SSO) on a Catalyst 6500 VSS, port load-share deferral is recommended on a port channel of a switch that is connected by a multichassis EtherChannel (MEC) to a VSS. Port load-share deferral is not supported on the Catalyst 4500/4948 and the Nexus 7000.

IS-IS Default Metric Style

On IOS systems, IS-IS must be configured to use transit or wide metric style to interoperate with the Nexus 7000.

IOS—Defaults to old style (narrow).

NX-OS—Defaults to new style (wide).

OSPF Interface Metric

The reference bandwidth for calculating OSPF metric is 100Mbps for IOS and 40Gbps for NX-OS. The entire OSPF network should be configured to use the same reference bandwidth. For NVT, 100Gbps was used.

BGP Per-Interface Fast External Failover

On NX-OS, per-interface fast failover applies only to eBGP peers. On IOS, it applies to all BGP peer types.

Multicast Multipath

IOS—Multicast multipath is disabled by default. When multipath is enabled, the default load sharing selection algorithm is source-based. The algorithm on IOS can be configured to match the behavior on NX-OS with the following command:

```
ip multicast multipath s-g-hash basic
```

NX-OS—Multicast multipath is enabled by default and the load sharing selection algorithm is based on the source and group addresses.

Multicast SPT Threshold

IOS—Multicast group filtering for spt-threshold is configured using an IP access list.

NX-OS—Multicast group filtering for spt-threshold is configured using a route-map. Within the route-map, the group filter can be specified using prefix address and mask; the group-range command, though available, is not supported.

Dynamic Trunking Protocol

Dynamic Trunking Protocol (DTP) is not supported in NX-OS. Configure the trunk port for unconditional trunking on the Catalyst 6500 when interoperating with the Nexus 7000.

Conclusion

By following the recommendations and guidelines suggested in this document, customers who deploy Cisco Catalyst 6500 and Nexus 7000 Series Switches can expect that the hardware and software features used in this network will behave and perform similarly between the platforms. Since NVT has been established as an additional quality assurance stage in order to leverage customer feedback and requirements into the product development cycle, future phases of NVT will continue to validate and publish additional guidelines for deploying Nexus 7000 and NX-OS solutions for datacenter networks. This document is intended to supplement the Cisco Nexus 7000 Series Switches product documentation that is available on cisco.com and should not be used as a replacement for that documentation.

GLOSSARY

A | B | C | D | E | F | G | H | I | K | L | M | N | O | P | R | S | U | V

A

ABR — See area border router.

ACE — access control entry.

ACL — access control list.

address family — A specific type of network addressing supported by a routing protocol. Examples include IPv4 unicast and IPv4 multicast.

adjacency — Two OSPF routers that have compatible configurations and have synchronized their link-state databases.

administrative distance — A rating of the trustworthiness of a routing information source. In general, the higher the value, the lower the trust rating.

area — A logical division of routers and links within an OSPF domain that creates separate subdomains. LSA flooding is contained within an area.

area border router — A router that connects one OSPF area to another OSPF area.

ARP — Address resolution protocol. ARP discovers the MAC address for a known IPv4 address.

AS — See autonomous system.

ASBR — See autonomous system border router.

ASM — Any Source Multicast. ASM is a PIM tree building mode.

attributes — Properties of a route that are sent in BGP UPDATE messages. These attributes include the path to the advertised destination as well as configurable options that modify the best path selection process.

autonomous system — A network controlled by a single technical administration entity.

autonomous system border router — A router that connect a an OSPF autonomous system to an external autonomous system.

B

backup designated router See BDR.

bandwidth The available traffic capacity of a link.

BDR Backup designated router. An elected router in a multi-access OSPF network that acts as the backup if the designated router fails. All neighbors form adjacencies with the backup designated router (BDR) as well as the designated router.

BFD Bidirectional Forwarding Detection. BFD is a detection protocol that is designed to provide fast forwarding path failure detection times.

BGP Border Gateway Protocol. BGP is an interdomain or exterior gateway protocol.

BGP peer A remote BGP speaker that is an established neighbor of the local BGP speaker.

BGP speaker BGP-enabled router.

Bidir-PIM Bidirectional Protocol Independent Multicast. Bidir-PIM is a variant of the PIM suite of routing protocols for IP multicast and is an extension of the existing PIM sparse mode (PIM-SM) feature.

C

CE customer edge.

communication cost Measure of the operating cost to route over a link.

converged The point at which all routers in a network have identical routing information.

convergence See converged.

CoPP Control Plane Policing.

D

dead interval The time within which an OSPF router must receive a Hello packet from an OSPF neighbor. The dead interval is usually a multiple of the hello interval. If no Hello packet is received, the neighbor adjacency is removed.

default gateway A router to which all unroutable packets are sent. Also called the router of last resort.

delay The length of time required to move a packet from the source to the destination through the internetwork.

designated router See DR.

DHCP Dynamic Host Control Protocol.

distance vector Defines routes by distance (for example, the number of hops to the destination) and direction (for example, the next-hop router) and then broadcasts to the directly connected neighbor routers.

DNS client Domain Name System client. Communicates with DNS server to translate a hostname to an IP address.

DR Designated router. An elected router in a multi-access OSPF network that sends LSAs on behalf of all its adjacent neighbors. All neighbors establish adjacency with only the designated router and the backup designated router.

E

eBGP External Border Gateway Protocol (BGP). Operates between external systems.

EIGRP Enhanced Interior Gateway Protocol. A Cisco routing protocol that uses the Diffusing Update Algorithm to provide fast convergence and minimized bandwidth usage.

F

feasible distance The lowest calculated distance to a network destination in EIGRP. The feasibility distance is the sum of the advertised distance from a neighbor plus the cost of the link to that neighbor.

feasible successor Neighbors in EIGRP that advertise a shorter distance to the destination than the current feasibility distance.

FHRP First Hop Redundancy Protocol.

FIB Fowarding Information Base. The forwarding table on each module that is used to make the Layer 3 forwarding decisions per packet.

FNF Flexible NetFlow.

G

gateway A switch or router that forwards Layer 3 traffic from a LAN to the rest of the network.

GLBP Gateway Load Balancing Protocol. A Cisco proprietary protocol that provides high availability features to end hosts.

graceful restart A feature that allows a router to remain in the data forwarding path while a routing protocol reboots.

GRE Generic Routing Encapsulation. A tunneling protocol that can encapsulate a wide variety of protocol packet types inside IP tunnels.

H

hello interval The configurable time between each Hello packet sent by an OSPF or EIGRP router.

hello packet A special message used by OSPF or IS-IS to discover neighbors. Also acts as a keepalive messages between established neighbors.

high availability The ability of a system or component to limit or avoid network disruption when a component fails.

hold time In BGP, the maximum time limit allowed in BGP between update or keepalive messages. If this time is exceeded, the TCP connection between the BGP peers is closed.

In EIGRP, the maximum time allowed between EIGRP Hello messages. If this time is exceeded, the neighbor is declared unreachable.

hop count The number of routers that can be traversed in a route. Used by RIP.

HSRP Hot Standby Router Protocol.

I

iBGP Internal Border Gateway Protocol (BGP). Operates within an autonomous system.

ICMP Internet Control Message Protocol.

IETF RFCs Internet Engineering Task Force Request for Comments.

IGMP Internet Group Management Protocol

IGP Interior Gateway Protocol. Used between routers within the same autonomous system.

instance An independent, configurable entity, typically a protocol.

IP tunnels A method of encapsulating packets within various Internet Protocols (IP) to interconnect communications between different networks.

IPv4 Internet Protocol version 4.

IPv6 Internet Protocol version 6.

IS-IS Intermediate System to Intermediate System. An ISO interior gateway protocol.

ISSU In-Service Software Upgrade.

K

keepalive A special message sent between routing peers to verify and maintain communications between the pair.

L

LACP Link Aggregation Control Protocol.

LDP MPLS Label Distribution Protocol.

link cost An arbitrary number configured on an OSPF interface which is in shortest path first calculations.

link-state Shares information about a link and link cost to neighboring routers.

link-state advertisement See LSA.

LSA Link-state advertisement. An OSPF message to share information on the operational state of a link, link cost, and other OSPF neighbor information.

link-state database OSPF database of all LSAs received. OSPF uses this database to calculate the best path to each destination in the network.

link-state refresh The time that OSPF floods the network with LSAs to ensure all OSPF routers have the same information.

load The degree to which a network resource, such as a router, is busy.

load balancing The distribution of network traffic across multiple paths to a given destination.

M

MD5 authentication digest A cryptographic construction that is calculated based on an authentication key and the original message and sent along with the message to the destination. Allows the destination to determine the authenticity of the sender and guarantees that the message has not been tampered with during transmission.

MEC multichassis EtherChannel.

message digest A one-way hash applied to a message using a shared password and appended to the message to authenticate the message and ensure the message has not been altered in transit.

metric A standard of measurement, such as the path bandwidth, that is used by routing algorithms to determine the optimal path to a destination.

MPLS Multi-Protocol Label Switching. MPLS is a packet-forwarding technology that uses labels to make data forwarding decisions.

MSDP Multicast Source Discovery Protocol

MTU Maximum transmission unit. The largest packet size that a network link transmits without fragmentation.

N

NDP Neighbor Discovery Protocol. The protocol used by IPv6 to find the MAC address associated with an IPv6 address.

NetFlow NetFlow is an embedded instrumentation within Cisco IOS software to characterize network operation.

network layer reachability information BGP network layer reachability information (NRLI). Contains the a list of network IP addresses and network masks for networks that are reachable from the advertising BGP peer.

next hop The next router that a packet is sent to on its way to the destination address.

NVT Nexus Validation Test.

O

OIR Online Insertion and Removal.

OSPF Open Shortest Path First. An IETF link-state protocol. OSPFv2 supports IPv4 and OSPFv3 supports IPv6.

P

path length Sum of all link costs or the hop count that a packet experiences when routed from the source to the destination.

PAgP Port Aggregation Protocol.

PIM Protocol Independent Multicast.

PIN Places in the Network. The Cisco PIN architecture addresses the differing requirements for systems design and deployment in the three principal network areas: the campus, the data center, Internet edge, and the Branch-WAN.

policy-based routing The method of using route maps to alter the route selected for a packet.

R

redistribution One routing protocol accepts route information from another routing protocol and advertises it in the local autonomous system.

Reliable Transport Protocol Responsible for guaranteed, ordered delivery of EIGRP packets to all neighbors.

reliability The dependability (usually described in terms of the bit-error rate) of each network link.

rendezvous point See RP.

RIB Routing Information Base. Maintains the routing table with directly connected routes, static routes, and routes learned from dynamic unicast routing protocols.

Route Policy Manager The process that controls route maps and policy-based routing.

Routing Information Base See RIB.

route map A construct used to map a route or packet based on match criteria and optionally alter the route or packet based on set criteria. Used in route redistribution and policy-based routing.

RP rendezvous point. An RP is a router in a multicast network domain that acts as a shared root for a multicast shared tree.

route summarization A process that replaces a series of related, specific routes in a route table with a more generic route.

router ID A unique identifier used by routing protocols. If not manually configured, the routing protocol selects the highest IP address configured on the system.

S

SPF algorithm Shortest Path First algorithm. Dijkstra's algorithm used by OSPF to determine the shortest route through a network to a particular destination.

split horizon Routes learned from an interface are not advertised back along the interface they were learned on, preventing the router from seeing its own route updates.

split horizon with poison reverse Routes learned from an interface are set as unreachable and advertised back along the interface they were learned on, preventing the router from seeing its own route updates.

SSM Source Specific Multicast. SSM is an extension of IP multicast where datagram traffic is forwarded to receivers from only those multicast sources to which the receivers have explicitly joined.

SSO/NSF Stateful Switchover with Nonstop Forwarding.

static route A manually configured route.

STP Spanning Tree Protocol.

stub area An OSPF area that does not allow AS External (type 5) LSAs.

stub router A router that has no direct connection to the main network and which routes to that network using a known remote router.

SVI switched virtual interface.

U

U6FIB Unicast IPv6 Forwarding Information Base.

UDLD Unidirectional Link Detection.

UFIB Unicast Forwarding Information Base for IPv4.

U6RIB Unicast IPv6 Routing Information Base. The unicast routing table that gathers information from all routing protocols and updates the forwarding information base for each module.

URIB Unicast Routing Information Base for IPv4. The unicast routing table that gathers information from all routing protocols and updates the forwarding information base for each module.

V

VDC virtual device context. Used to split a physical system into secure, independent, logical systems.

virtualization A method of making a physical entity act as multiple, independent logical entities.

vPC virtual PortChannel. A vPC allows links that are physically connected to two different devices to appear as a single PortChannel to a third device.

VRF virtual routing and forwarding. A method used to create separate, independent Layer 3 entities within a system, or an instance of that method.

VRF-lite VRF-lite (MPLS Multi-VRF) provides the ability to configure and maintain more than one instance of a routing and forwarding table within the same customer edge (CE) router.

VRRP Virtual Router Redundancy Protocol.

VSS virtual switching system. A VSS is network system virtualization technology that pools multiples witches into one virtual switch.