SPADVROUTE

# Deploying Cisco Service Provider Advanced Network Routing

## Student Guide

## Volume 2

Version 1.2

Part Number: 97-3434-01

https://t.me/learningnets




| Americas Headquarters | Asia Pacific Headquarters | Europe Headquarters |
|---|---|---|
| Cisco Systems, Inc.<br>San Jose, CA | Cisco Systems (USA) Pte. Ltd.<br>Singapore | Cisco Systems International BV<br>Amsterdam,<br>The Netherlands |

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at www.cisco.com/go/offices.



Cisco and the Cisco logo are trademarks or registered trademarks of Cisco and/or its affiliates in the U.S. and other countries. To view a list of Cisco trademarks, go to this URL: 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. (1110R)

DISCLAIMER WARRANTY: THIS CONTENT IS BEING PROVIDED "AS IS" AND AS SUCH MAY INCLUDE TYPOGRAPHICAL, GRAPHICS, OR FORMATTING ERRORS. CISCO MAKES AND YOU RECEIVE NO WARRANTIES IN CONNECTION WITH THE CONTENT PROVIDED HEREUNDER, EXPRESS, IMPLIED, STATUTORY OR IN ANY OTHER PROVISION OF THIS CONTENT OR COMMUNICATION BETWEEN CISCO AND YOU. CISCO SPECIFICALLY DISCLAIMS ALL IMPLIED WARRANTIES, INCLUDING WARRANTIES OF MERCHANTABILITY, NON-INFRINGEMENT AND FITNESS FOR A PARTICULAR PURPOSE, OR ARISING FROM A COURSE OF DEALING, USAGE OR TRADE PRACTICE. This learning product may contain early release content, and while Cisco believes it to be accurate, it falls subject to the disclaimer above.

© 2014 Cisco Systems, Inc.



https://t.me/learningnets

# Table of Contents


**Module 5: Intradomain and Interdomain Multicast Routing ..... 1**
Lesson 1: Introducing PIM-SM Protocol ..... 3
PIM-SM in the Cisco IP NGN Infrastructure Layer ..... 5
PIM-SM Shared Tree Join ..... 7
PIM-SM Sender Registration ..... 8
PIM-SM SPT Switchover ..... 9
PIM-SM Packets ..... 12
PIM-SM State Information ..... 14
PIM-SM State Maintenance ..... 15
Multicast Routing Table ..... 16
PIM-SM OIL Rules ..... 19
PIM-SM State Flags ..... 20
PIM-SM Neighbor Discovery ..... 22
PIM-SM Forwarding ..... 23
PIM-SM Joining ..... 24
PIM-SM Registering ..... 25
PIM-SM Registering: Receiver Joins First Scenario ..... 27
PIM-SM Registering: Source Starts First Scenario ..... 35
PIM-SM Registering: Receivers Along the SPT Scenario ..... 43
PIM-SM SPT Switchover Overview ..... 49
PIM-SM SPT Switchover Process ..... 53
PIM-SM Shared Tree Pruning Overview ..... 60
PIM-SM Pruning Shared Tree Process ..... 63
PIM-SM SPT Pruning Overview ..... 65
PIM-SM Pruning SPT Process ..... 67
Enable PIM-SM ..... 72
Static RP ..... 73
Finding PIM Neighbors and Checking RP Information ..... 74
Troubleshooting PIM-SM Guidelines ..... 76
Summary ..... 77
Lesson 2: Implementing PIM-SM Enhancements ..... 79
Source Specific Multicast ..... 81
SSM Scenario ..... 83
SSM with IGMPv3 ..... 84
SSM Mapping ..... 85
Configuring SSM ..... 86
Bidirectional PIM ..... 87
Bidirectional PIM Sources and Receivers ..... 89
Bidirectional PIM Traffic Flow ..... 91
Forwarding and Tree Building Process ..... 92
DF Election Process ..... 94
DF Election Messages ..... 95
Initial Election ..... 96
Winning the Election as DF ..... 97
DF Loses Path to the RP ..... 98
DF Dies and Other Metric Changes ..... 99
Configuring Bidirectional PIM ..... 101
Summary ..... 102





https://t.me/learningnets

Lesson 3: Implementing Interdomain IP Multicast ..... 103
Service Provider Multicast Requirements ..... 105
GLOP—Static Allocation of 233/8 ..... 106
SSM Role in Interdomain IP Multicast ..... 107
MSDP Role in Interdomain IP Multicast ..... 109
Multicast Service Provider Requirements ..... 110
MSDP Protocol ..... 111
MSDP Concepts ..... 113
MSDP Neighbor Relationship ..... 117
MSDP Messages ..... 118
MSDP SA Message Processing ..... 119
MSDP SA Message Origination ..... 120
MSDP MD5 Password Authentication ..... 122
Configuring MSDP ..... 123
Verifying MSDP ..... 124
Summary ..... 126
Lesson 4: Identifying Rendezvous Point Distribution Solutions ..... 127
Static RP Disadvantages ..... 129
Dynamic RP Discovery Mechanisms ..... 130
RP Placement ..... 131
Auto-RP ..... 132
Auto-RP Candidate RPs ..... 133
Auto-RP Mapping Agents ..... 134
Auto-RP Other Routers ..... 135
Auto-RP Configuration ..... 136
Auto-RP Troubleshooting ..... 138
Auto-RP Scoping ..... 142
Securing Auto-RP Using a Boundary ..... 144
PIMv2 Bootstrap Router ..... 146
PIMv2 BSR Candidate RPs ..... 148
PIMv2 BSRs ..... 149
PIMv2 BSR Election ..... 150
Other PIMv2 BSR Routers ..... 152
PIMv2 BSR Advertisement Process ..... 153
PIMv2 BSR Configuration ..... 154
PIMv2 BSR Verification ..... 155
PIMv2 BSR Troubleshooting ..... 156
BSR Hop-by-Hop Flooding ..... 157
Constraining BSR Messages ..... 158
Anycast RP ..... 159
Anycast RP Example ..... 161
Anycast RP Configuration ..... 163
Anycast RP Configuration Guidelines ..... 164
Summary ..... 165
Lesson 5: Module Summary ..... 167
Lesson 6: Module Self-Check ..... 169
Module Self-Check Answers ..... 177
Answer Key ..... 177
**Module 6: Service Provider IPv6 Transition Implementations ..... 179**
Lesson 1: Introducing IPv6 Services ..... 181
Multicast in the Cisco IP NGN Infrastructure Layer ..... 183
IPv6 Multicast Address Format ..... 184
IPv6 Multicast Address Scope ..... 186



https://t.me/learningnets

IPv6 Solicited-Node Multicast Address Format ..... 188
IPv6 Multicast Address with a Global Scope ..... 190
IPv4 vs. IPv6 Multicast Comparison ..... 191
PIMv6 Overview ..... 192
Embedding the RP Address in an IPv6 Multicast Address ..... 197
IPv6 Multicast Routing Configuration ..... 200
IPv6 Multicast Listener Discovery ..... 201
MLDv1 Messages ..... 204
MLDv1 General Query Message ..... 205
MLDv1 Report Message ..... 206
MLDv1 Done Message ..... 208
MLDv1 Address-Specific Query Message ..... 209
MLDv2 Protocol ..... 210
MLD Access Groups and Group Limits ..... 212
MLD Configuration ..... 213
MLD Verification ..... 214
MLD Join-Group and Static-Group ..... 215
MLD Snooping ..... 216
DNS and DHCPv6 in the Cisco IP NGN Infrastructure Layer ..... 219
DNS IPv6 Support ..... 220
Dynamic DNS ..... 226
DHCPv6 Operations ..... 228
DHCPv6 Server Router Configuration ..... 230
DHCPv6 Lite Operation (Stateless DHCPv6) ..... 232
DHCPv6 Prefix Delegation ..... 233
DHCPv6 Verification ..... 237
QoS in the Cisco IP NGN Infrastructure Layer ..... 239
IPv6 Header Fields Used for QoS ..... 240
IPv6 Traffic Class Field ..... 241
IPv6 Flow Label Field ..... 243
IPv6 QoS Configuration ..... 246
Cisco IOS Software Features ..... 248
Cisco IOS IPv6 Telnet and SSH Server and Client Support ..... 249
Cisco IOS IPv6 Tools ..... 251
Cisco Discovery Protocol Support for IPv6 ..... 253
Cisco Express Forwarding for IPv6 ..... 255
IP Service Level Agreement (SLA) for IPv6 ..... 256
Configuring IP SLA ..... 258
Summary ..... 261
Lesson 2: Defining IPv6 Transition Mechanisms ..... 265
Dual Stack, CGN, and NAT64 ..... 267
Dual-Stack Operations Overview ..... 268
Dual-Stack Considerations ..... 272
Dual Stack with Carrier-Grade NAT ..... 273
NAT444 ..... 275
Carrier-Grade NAT on Cisco Routers ..... 276
NAT64 ..... 278
DNS64 ..... 279
Stateless NAT64 ..... 281
Stateful NAT64 ..... 283
Stateless versus Stateful NAT64 Comparison ..... 284
Stateful NAT64 Configuration on ASR 1000 ..... 285
Static, Stateful NAT64 Configuration on ASR 1000 ..... 286
IPv6 Tunneling Mechanisms ..... 289



https://t.me/learningnets

Manually Configured Tunnels ..... 292
GRE Tunnels ..... 293
6in4 Tunnels ..... 294
6in4 Tunnel Configuration ..... 295
6to4 Automatic Tunnels ..... 297
6to4 Considerations ..... 299
6to4 Configuration ..... 300
6RD Automatic Tunnels ..... 302
6RD Considerations ..... 305
6RD Configuration ..... 306
Summary ..... 308
Lesson 3: Deploying IPv6 in the Service Provider Network ..... 311
IPv6 Service Provider Deployment ..... 313
Dual-Stack Option ..... 315
Dual Stack Pros and Cons ..... 316
Tunneling of IPv6 in IPv4 ..... 318
Tunneling of IPv6 in IPv4 Pros and Cons ..... 319
Key Service Provider Strategies ..... 320
IPv6 Services ..... 321
IPv6 Address Allocation ..... 323
IPv6 Address Selection Guidelines ..... 325
IPv6 Broadband Access Services ..... 326
FTTH Access Architecture ..... 328
DSL Access Architecture ..... 329
Cable Access Architecture ..... 330
Summary ..... 331
Lesson 4: Module Summary ..... 333
Lesson 5: Module Self-Check ..... 335
Module Self-Check Answers ..... 337
Answer Key ..... 337
**Glossary ..... 339**

 © 2014 Cisco Systems, Inc.

https://t.me/learningnets

# Module 5: Intradomain and Interdomain Multicast Routing

## Introduction

This module provides a detailed explanation of the most current scalable IP multicast routing protocol, PIM-SM. This explanation includes the principles of operation and protocol details. Learners will become familiar with the determinism built into sparse mode multicast protocols, and will be prepared for complex deployments of IP multicast. This module offers an overview of RP distribution solutions.

Interdomain IP multicast forwarding is constrained by the lack of proper standardized protocols for building interdomain distribution trees. MSDP is a solution that allows interdomain multicast routing by announcing sources from other domains into local domains. This capability interconnects multicast distribution trees. This module also explains the basic concepts of MSDP and its use in the IP multicast environment.

This module explains the issues and challenges with IP multicast security and describes the process of monitoring and maintaining multicast high-availability operations.

Upon completing of this module, you will be able to:

- Describe the principles of operation of PIM-SM and explain various control mechanisms in maintaining the distribution tree
- Describe SSM, PIM-SM bidirectional mode, and IGMPv3
- Implement MSDP in the interdomain environment
- Introduce the need for dynamic RP information distribution, and list mechanisms for dynamic RP distribution



https://t.me/learningnets

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# Lesson 1: Introducing PIM-SM Protocol

## Overview

This lesson provides a detailed explanation of the most current scalable IP multicast routing protocol, PIM-SM. This explanation includes the principles of operation and protocol details. Learners will become familiar with the determinism built into sparse mode multicast protocols and will become prepared for complex deployments of IP multicast.

This lesson describes the principles and operation of PIM-SM and the contents of the protocol packets. Major steps in PIM-SM operation are described, and various control mechanisms in maintaining the distribution tree are explained.

Upon completing this lesson, you will be able to:

- Explain the principles and operation of PIM-SM
- Discuss the PIM-SM shared-tree join process
- Discuss the PIM-SM sender registration process
- Discuss the PIM-SM SPT switchover process
- Discuss the various PIM-SM control packets
- Discuss PIM-SM state information
- Discuss PIM state maintenance
- Show examples of the multicast routing table and discuss PIM-SM (*,G) and PIM-SM (S, G) state rules
- Discuss PIM-SM OIL rules
- Discuss PIM-SM state flags
- Describe the operation of PIM-SM neighbor discovery
- Describe the operation of PIM-SM forwarding



https://t.me/learningnets

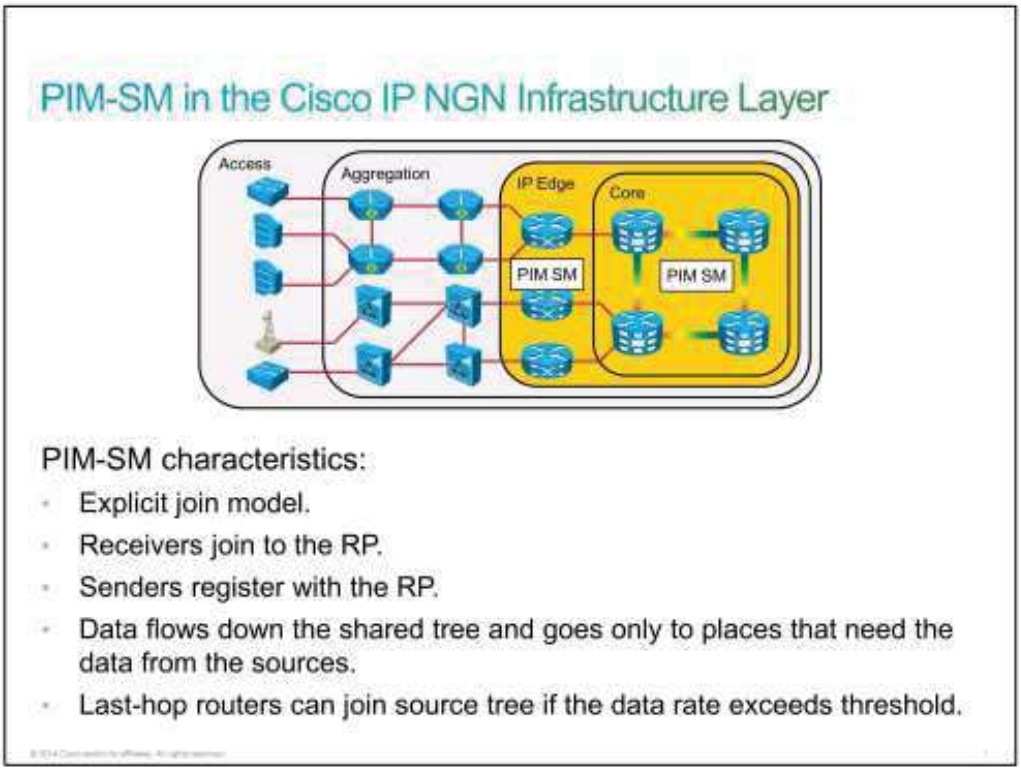
- Describe the operation of PIM-SM joining
- Describe the operation of PIM-SM registering
- Explain the register process when the shared tree exists, meaning that the receivers joined the group before the source started sending multicast traffic
- Explain the register process when the source starts but there are no receivers
- Explain the register process when there are receivers along the SPT between the first-hop router and the RP
- Provide an overview of the PIM-SM SPT switchover process
- Explain the PIM-SM SPT switchover process
- Provide an overview of the PIM-SM shared-tree pruning process
- Explain the PIM-SM shared-tree pruning process
- Provide an overview of the PIM-SM SPT pruning process
- Explain the PIM-SM SPT pruning process
- Explain how to enable PIM-SM
- Explain implementation of PIM-SM using static RP
- Explain how to find PIM neighbors and check RP information
- Explain PIM-SM troubleshooting guidelines

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# PIM-SM in the Cisco IP NGN Infrastructure Layer

PIM-SM uses the explicit join model and is used in the IP edge and core layers of the Cisco IP Next-Generation Network. Receivers send PIM join messages to a designated RP. The RP is the root of a shared distribution tree down which all multicast traffic flows.

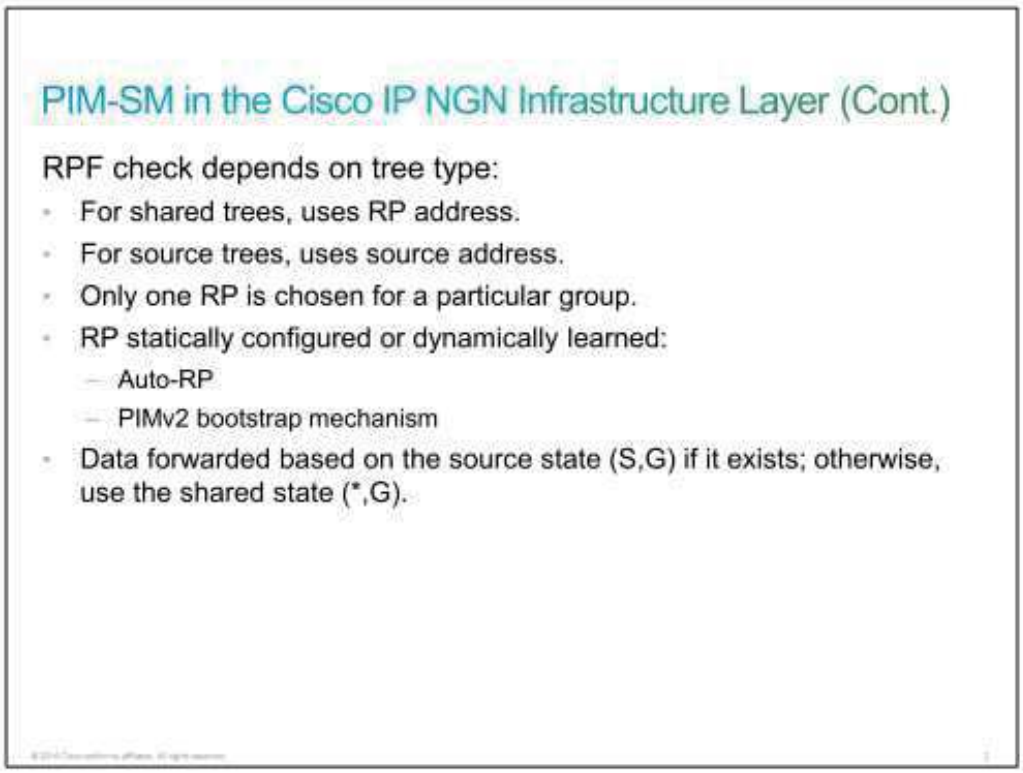



To get multicast traffic to the RP for distribution down the shared tree, first-hop routers with directly connected senders send PIM register messages to the RP. Register messages cause the RP to send an (S,G) join toward the source. This activity enables multicast traffic to flow natively to the RP via a SPT, and then down the shared tree.

Routers may be configured with an SPT threshold, which, once exceeded, will cause the last-hop router to join the SPT. This action will cause the multicast traffic from the source to flow down the SPT directly to the last-hop router.



https://t.me/learningnets




The RPF check is done differently, depending on tree type. If traffic is flowing down the shared tree, the RPF check mechanism will use the IP address of the RP to perform the RPF check. If traffic is flowing down the SPT, the RPF check mechanism will use the IP address of the source to perform the RPF check.

Although it is common for a single RP to serve all groups, it is possible to configure different RPs for different groups or group ranges. This approach is accomplished via access lists. Access lists permit you to place the RPs in different locations in the network for different group ranges. The advantage to this approach is that it may improve or optimize the traffic flow for the different groups. However, only one RP for a group may be active at a time.

RPs may be configured statically on each router. If you perform static configuration, you must ensure that all routers in the network agree on the RPs, otherwise the network will be broken.

However, a better solution is to use one of the following mechanisms, which automatically distribute the RP information:

- Auto-RP, which is the Cisco implementation
- Standard PIMv2 bootstrap mechanism, by using BSRs

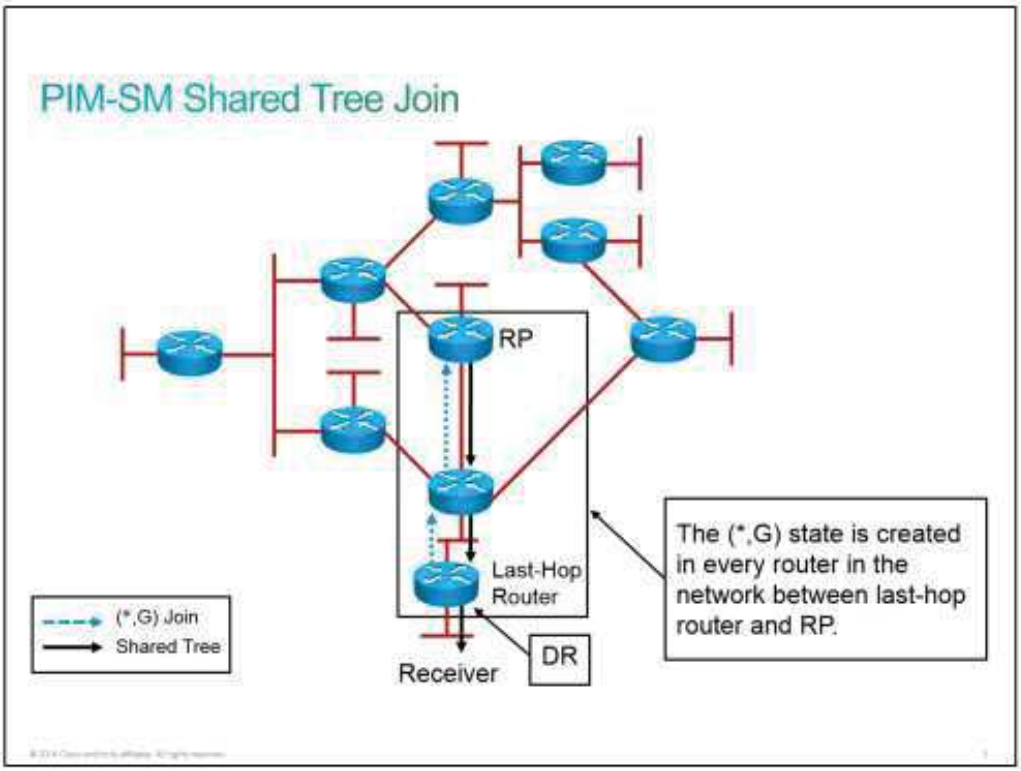
Multicast traffic forwarding in a PIM-SM network is first attempted using any matching (S,G) entries in the MRIB. If no matching (S,G) state exists, then the traffic is forwarded using the matching (*,G) entry in the multicast routing table.



https://t.me/learningnets

# PIM-SM Shared Tree Join

In this figure, an active receiver has joined multicast group G by multicasting an IGMP membership report. A DR on the LAN segment will receive IGMP membership reports.




The DR knows the IPaddress of the RP router for group G and sends a (*,G) join for this group toward the RP.

This (*,G) join travels hop by hop toward the RP, building a branch of the shared tree that extends from the RP to the last-hop router directly connected to the receiver.

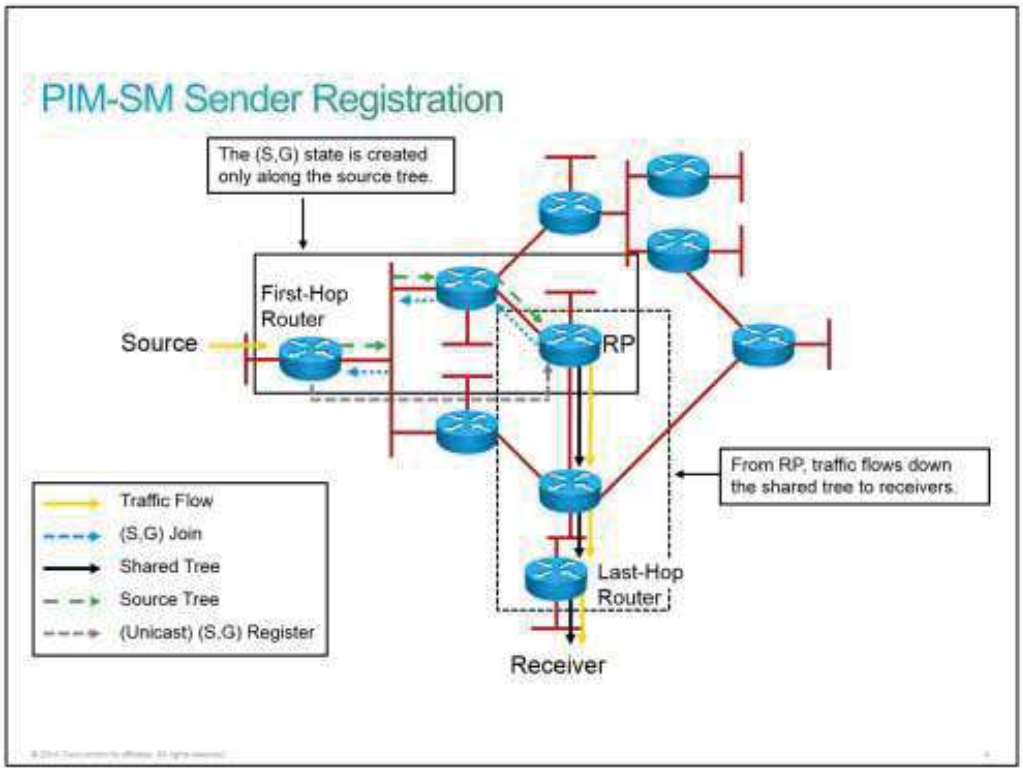
At this point, group G traffic may flow down the shared tree to the receiver.



https://t.me/learningnets

# PIM-SM Sender Registration

As soon as an active source for group G starts sending multicast packets, its first-hop DR registers the source with the RP. To register a source, the DR encapsulates the multicast packets in a PIM register message and sends the message to the RP using unicast.




When the RP receives the unicast register message, it de-encapsulates the multicast packets inside the unicast register message and starts sending the multicast packets down the shared tree toward the receivers. At the same time, the RP initiates the building of an SPT from the source to the RP by sending (S,G) joins toward the source. The building of an SPT causes the creation of an (S,G) state in all routers along the SPT, including the RP.

After the SPT is built from the first-hop router to the RP, the multicast traffic starts to flow from the source (S) to the RP without being encapsulated in the unicast register messages.

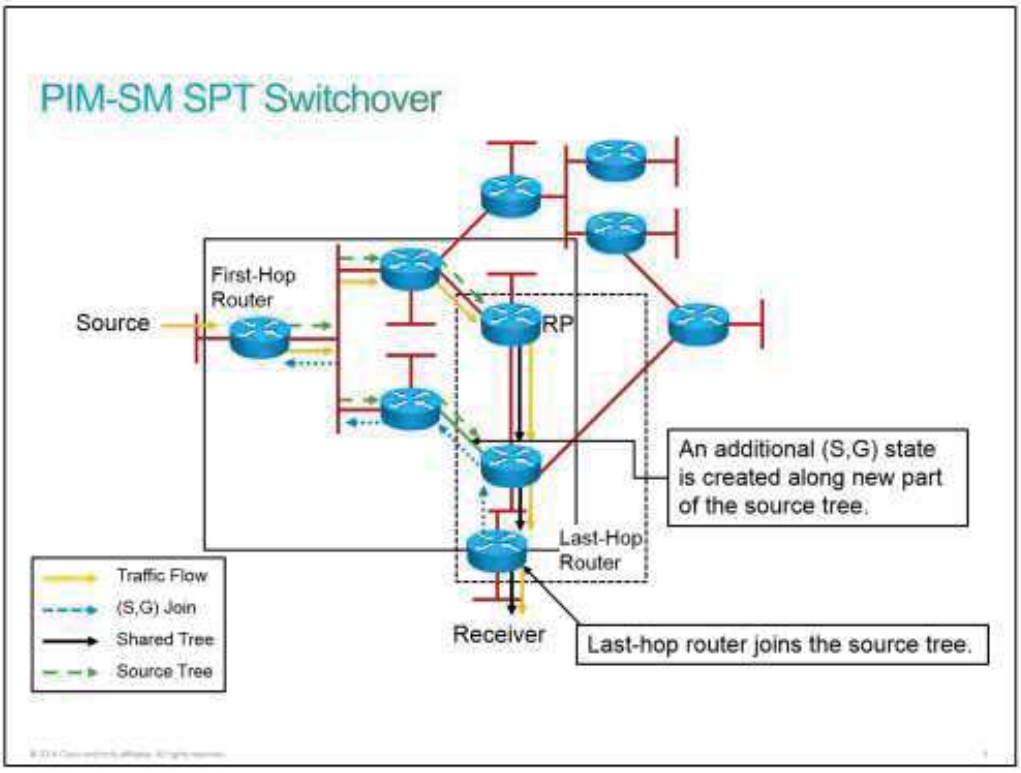
When the RP begins receiving multicast data down the SPT from the source, it sends a unicast PIM register-stop message to the first-hop router. The PIM register-stop message informs the first-hop router that it may stop sending the unicast register messages. At this point, the multicast traffic from the source is flowing down the SPT to the RP and, from there, down the shared tree to the receivers.

 © 2014 Cisco Systems, Inc.

https://t.me/learningnets

# PIM-SM SPT Switchover

PIM-SM enables the last-hop router (the router with directly connected receivers) to switch to the SPT and bypass the RP if the multicast traffic rate is above a set threshold. This threshold is called the SPT threshold.



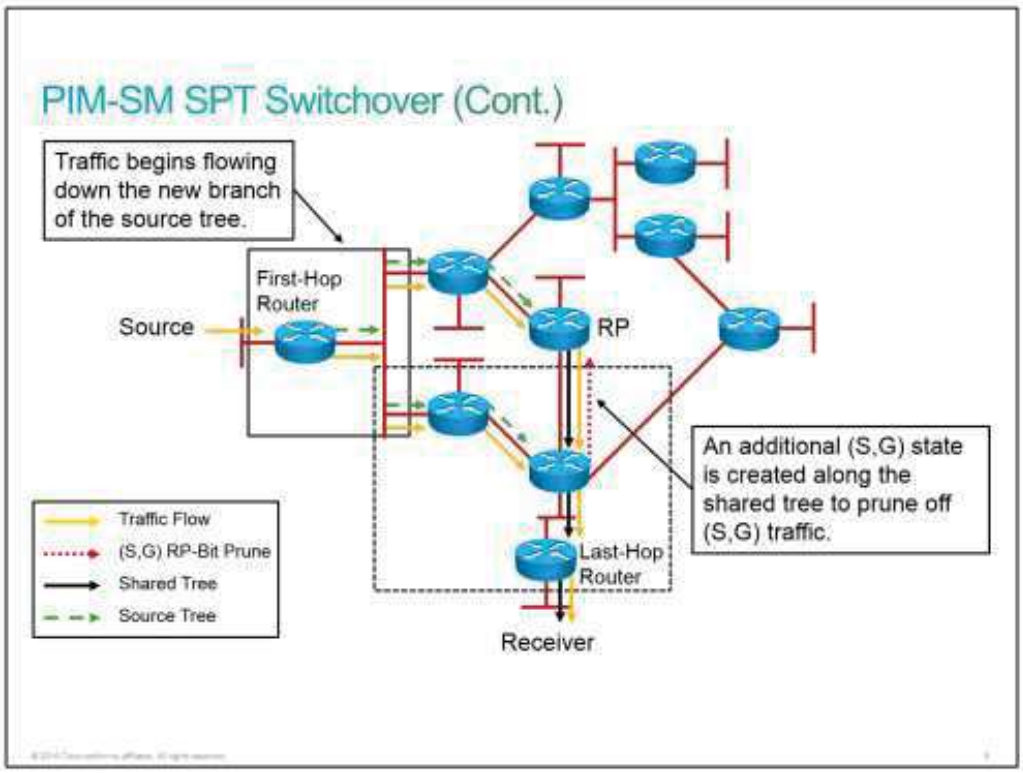

In Cisco routers, the default value of the SPT threshold is 0. Thus, as soon as the first packet arrives via the (*,G) shared tree, the default action for Cisco PIM-SM routers that are attached to active receivers is to immediately join the SPT to the source.

In this figure, the last-hop router (at the bottom of the figure) sends an (S,G) join message toward the source to join the SPT and bypass the RP.

This (S,G) join message travels hop by hop to the first-hop router (the router that is connected directly to the source), thereby creating another branch of the SPT. This action also creates an (S,G) state in all routers along this branch of the SPT.

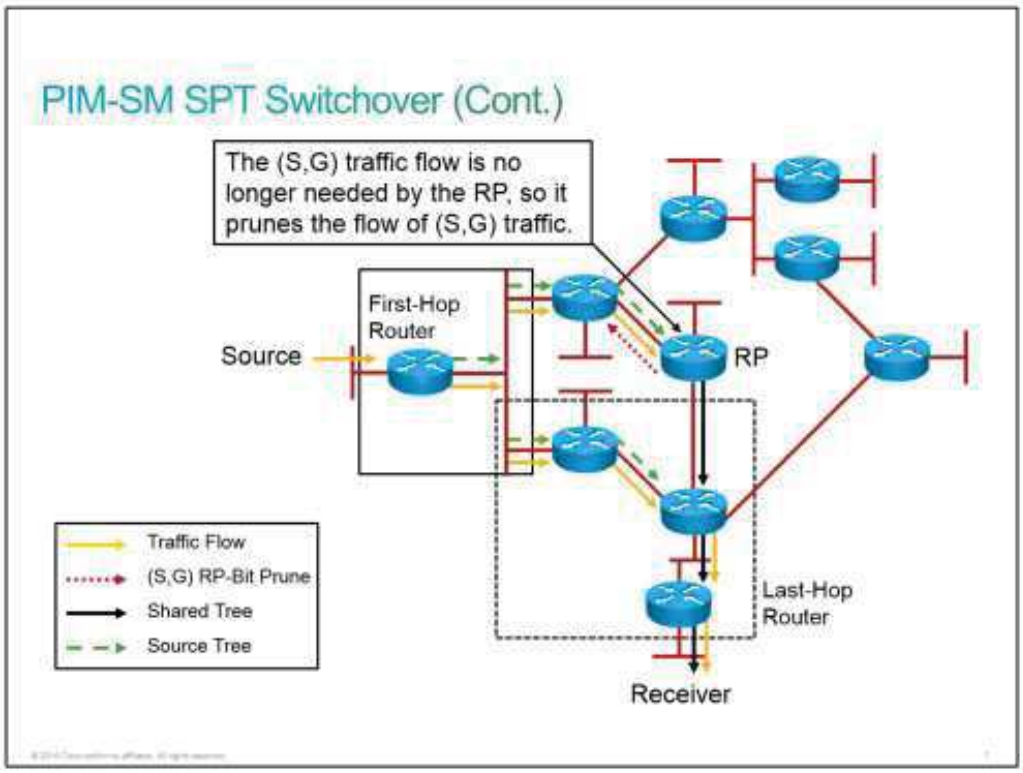


https://t.me/learningnets




(S,G) traffic is now flowing directly from the source to the receiver via the SPT. A special (S,G) RP-bit prune message is sent up the shared tree to prune the (S,G) traffic from the shared tree. If the prune message were not sent, duplicate (S,G) traffic would continue flowing down the shared tree to the receiver.

After the (S,G) RP-bit prune message has reached the RP, the branch of the SPT from the source to the RP still exists. However, when the RP has received the (S,G) RP-bit prune message via all branches of the shared tree, the RP no longer needs this (S,G) traffic. This result occurs because all of the receivers in the network are receiving the traffic via an SPT, bypassing the RP.




 © 2014 Cisco Systems, Inc.

https://t.me/learningnets

The RP no longer needs the (S,G) traffic. Therefore, it will send (S,G) prune messages back toward the source to shut off the flow of now unnecessary (S,G) traffic.

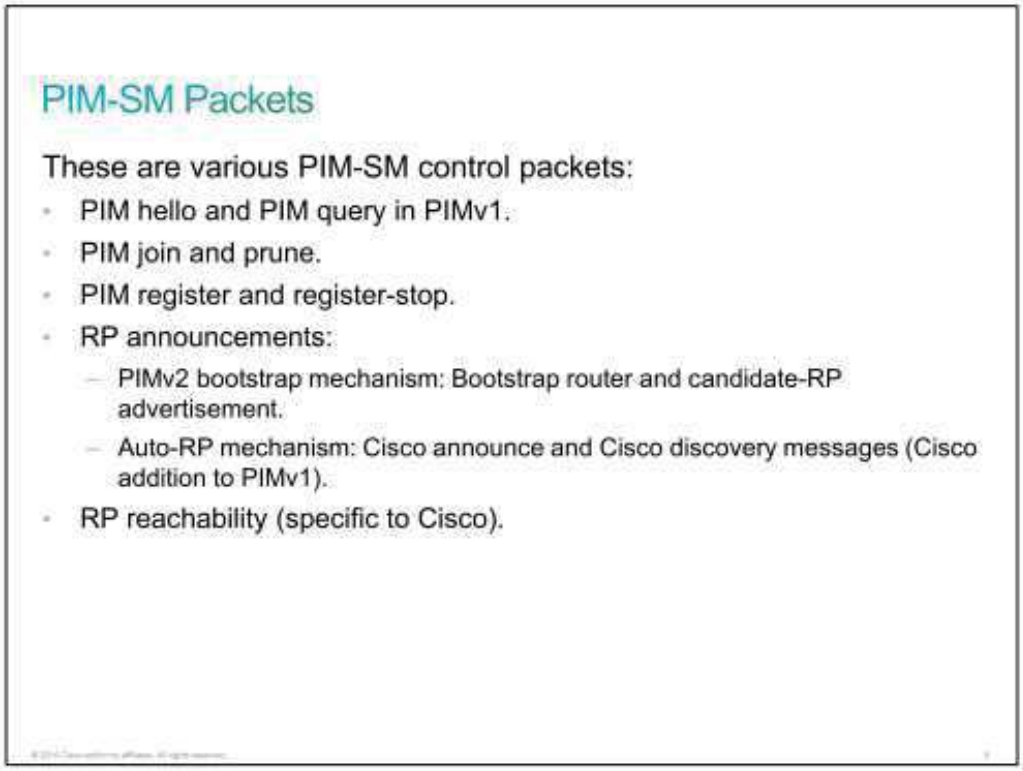
When the (S,G) prune message reaches the first-hop router, it prunes the branch of the (S,G), thus tearing down the SPT that was built between the source and the RP. The traffic is now only flowing down the remaining branch of the (S,G)—the SPT built between the source and the last-hop router that initiated the switchover. Because of the SPT switchover mechanism, PIM-SM also supports the construction and use of SPT (S,G) trees, but in a more economical way than the PIM-DM forwarding state.



https://t.me/learningnets

# PIM-SM Packets

The PIM-SM has various control packets.




PIM-SM for PIMv2 packets are encapsulated into IP packets with protocol number 103. PIMv1 uses IGMP packets. PIM-SM has several different message types, as follows:

- PIM hello message (formerly PIMv1 query message) for discovering PIM neighbors and maintaining neighbor adjacencies. These hello packets are sent periodically (every 30 seconds, by default) via multicast address 224.0.0.13 to indicate to the other PIM routers on the network that this PIM router is still present. In addition, the hold time indicates how long a router will keep a neighbor that is no longer sending messages, and an option is available in the options field that assists in electing the DR. The DR priority may be set to force a router on a multiaccess segment to become the DR.
- PIM join and prune messages for joining and pruning multicast traffic. The join or prune packet is a single-packet format that contains both a list of joins and a list of prunes. Either list—but not both—may be empty. Joins are primarily used in PIM-SM, which is based on the explicit join model. Prune messages are used to cut off the traffic from the upstream router if there are no receivers downstream.
- PIM register and register-stop messages for registering multicast sources. The designated first-hop router sends the unicast PIM register message to the RP when the source starts multicast traffic. A PIM register-stop message is sent from the RP toward the sender via unicast. After the first-hop router receives this message, it stops encapsulating multicast packets in the PIM register message and starts sending them via multicast. The register-stop message is sent as follows:
  - Soon after the first packet arrives natively down the SPT built between the RP and the first-hop router
  - Immediately if there is no shared tree for the group for which the first-hop router is registering to the RP
- Candidate-RP messages, which send PIM candidate-RP advertisements (or Cisco announce messages):



https://t.me/learningnets

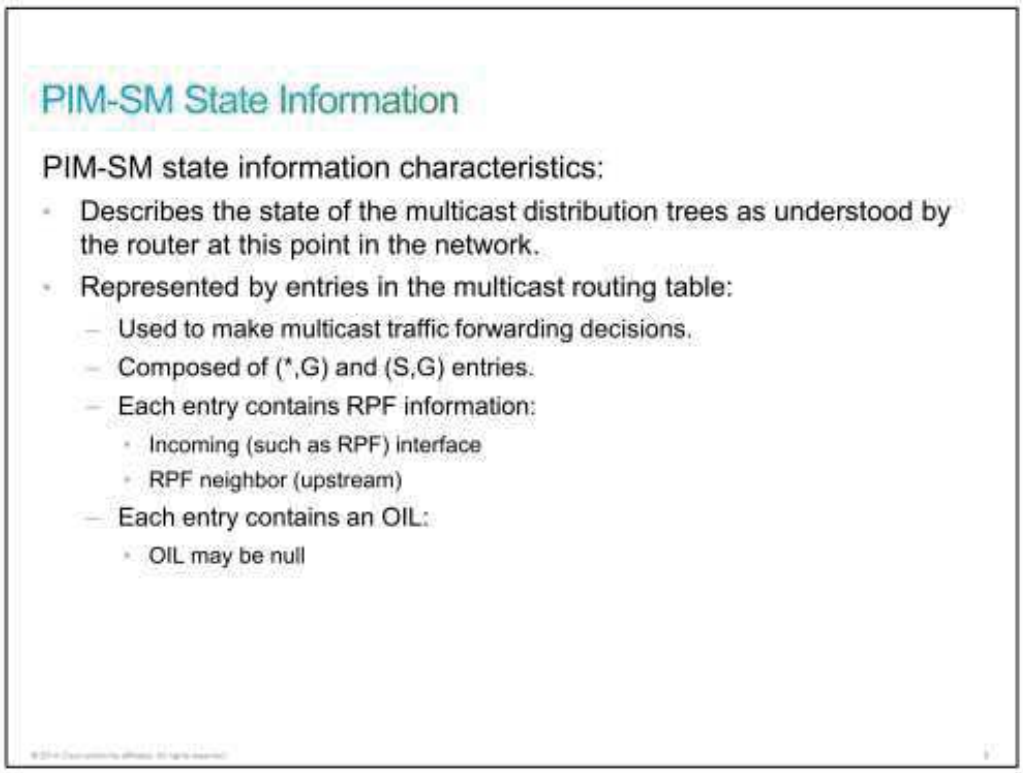
  - The BSR in PIMv2 collects RP information and distributes it in PIM bootstrap messages. PIM bootstrap messages are also used to dynamically elect the BSR.
  - The PIM Auto-RP mechanism is the Cisco implementation for automated distribution of group-to-RP mappings in a PIM-SM network. The function, similar to BSR, is assigned to a mapping agent in Auto-RP—the mapping agent is responsible for sending Cisco discovery messages to the Cisco Discovery (224.0.1.40) multicast group.
- The PIM RP reachability message (specific to Cisco), which is used by the RP to report its reachability to all the nodes on a shared tree.



https://t.me/learningnets

# PIM-SM State Information

In general, the multicast state describes the multicast distribution trees as they are understood by the router at a particular point in the network. However, the multicast state also describes the multicast-traffic-forwarding state that is used by the router to forward multicast traffic.




The multicast state is stored in the multicast routing table, which may be displayed using the Cisco IOS and IOS XE **show ip mroute** command or the Cisco IOS XR **show mrib ipv4 route** command.

Entries in the multicast route table are composed of (*,G) and (S,G) entries, each of which contains the following:

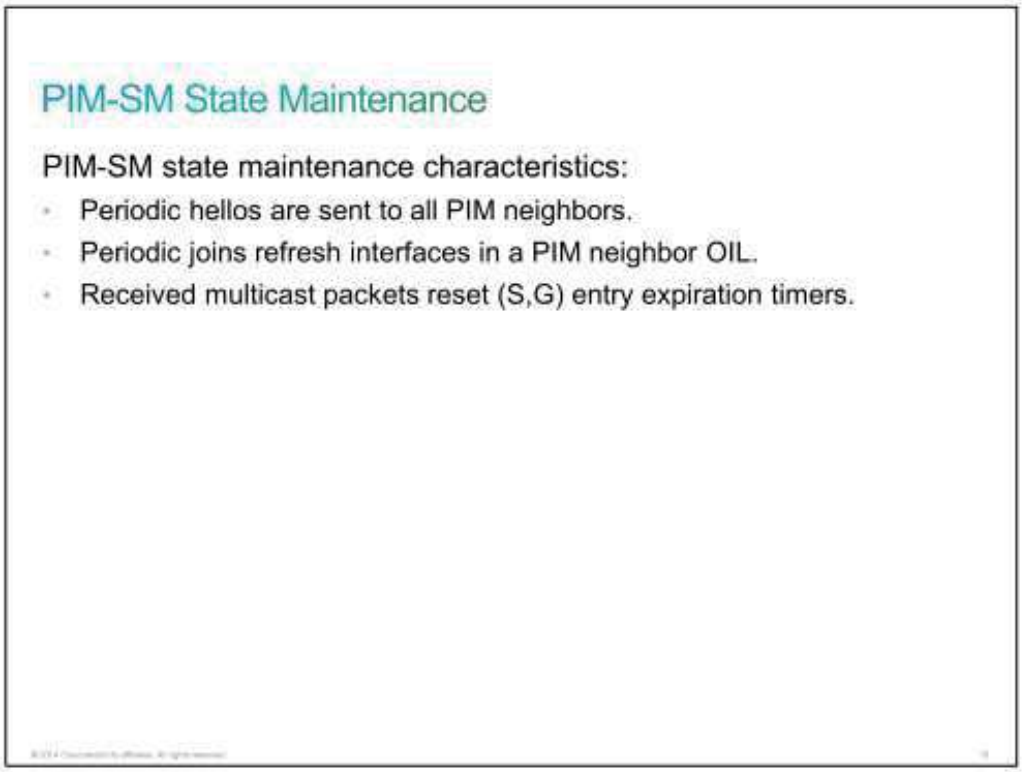
- RPF information, consisting of an RPF interface or an IIF and the IP address of the upstream RPF neighbor router in the direction of the source. With PIM-SM, the information in a (*,G) entry points toward the RP.
- OIL, which contains a list of interfaces over which the multicast traffic is going to be forwarded. Multicast traffic must arrive on the IIF (RPF interface) before it can be forwarded to these OIL interfaces. If multicast traffic does not arrive on the IIF, it is simply discarded.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# PIM-SM State Maintenance

In PIM-SM, join and prune state information has a normal expiration time of 3 minutes. If a periodic join and prune message is not received to refresh this state information, it automatically expires and is deleted. Therefore, a PIM router sends periodic join and prune messages to its PIM neighbors to maintain this state information.




When a join message is received from a PIM neighbor, the expiration timer of the interface (in the OIL) is reset to 3 minutes. If the interface expiration timer counts down to 0, the interface is removed from the OIL.

This action may trigger a prune if the removal of the interface causes the OIL to become null (empty).

When a prune message is received in PIM sparse mode, the interface on which the prune was received is normally just removed from the OIL. The exception is the special case of (S,G) RP-bit prunes, which are used to prune (S,G) traffic from the shared tree. In this case, periodic (S,G) RP-bit prunes must be sent to refresh the prune state in the upstream PIM neighbor toward the RP.

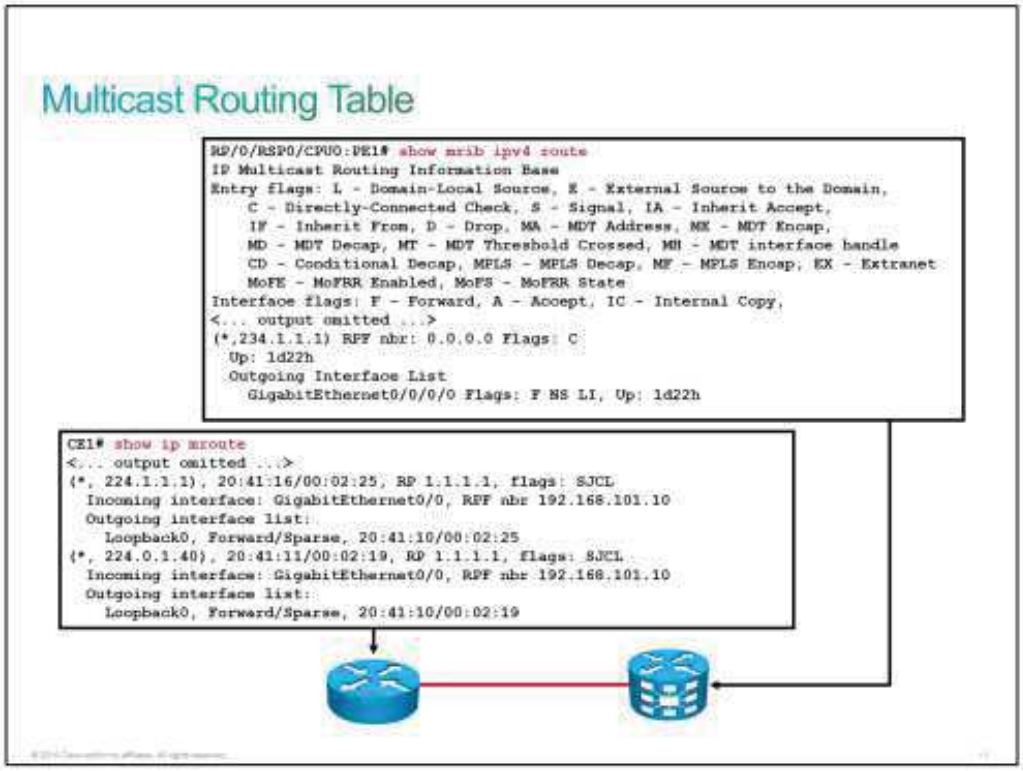
All (S,G) entries have entry expiration timers, which are reset to 3 minutes by the receipt of an (S,G) packet via the SPT. If the source stops sending, this expiration timer counts down to 0, and the (S,G) entry is deleted.



https://t.me/learningnets

# Multicast Routing Table

This figure shows the multicast routing table example with (*,G) entry.




The (*, 234.1.1.1) entry that is shown in the sample output is the (*,G) entry. If there is no matching entry for a particular (S,G) entry, this entry is used to forward the 234.1.1.1 multicast group traffic down the shared tree out the Gi0/0/0/0 interface.

## PIM-SM (*,G) State Rules

The PIM-SM (*,G) state rules are described below.

The following are the PIM-SM (*,G) state rules:

- The (*,G) entry is created when a (*,G) join or an IGMP report is received. The latter condition may be simulated by manually configuring the router interface to join the group.
- (*,G) entries are also created automatically whenever an (S,G) entry for the group must be created. The (*,G) entry is created first; then the (S,G) entry is created.
- The IIF reflects the RPF interface and neighbor in the direction of the RP.
- The OIL of a PIM-SM (*,G) entry reflects interfaces with one of these attributes:
  - Have received a (*,G) join.
  - Are directly connected to a member that has joined the group.
  - Have been manually configured to join the group. (This action may be accomplished by using the Cisco IOS and IOS XE **ip igmp static-group** command or the Cisco IOS XR **static-group** command.)

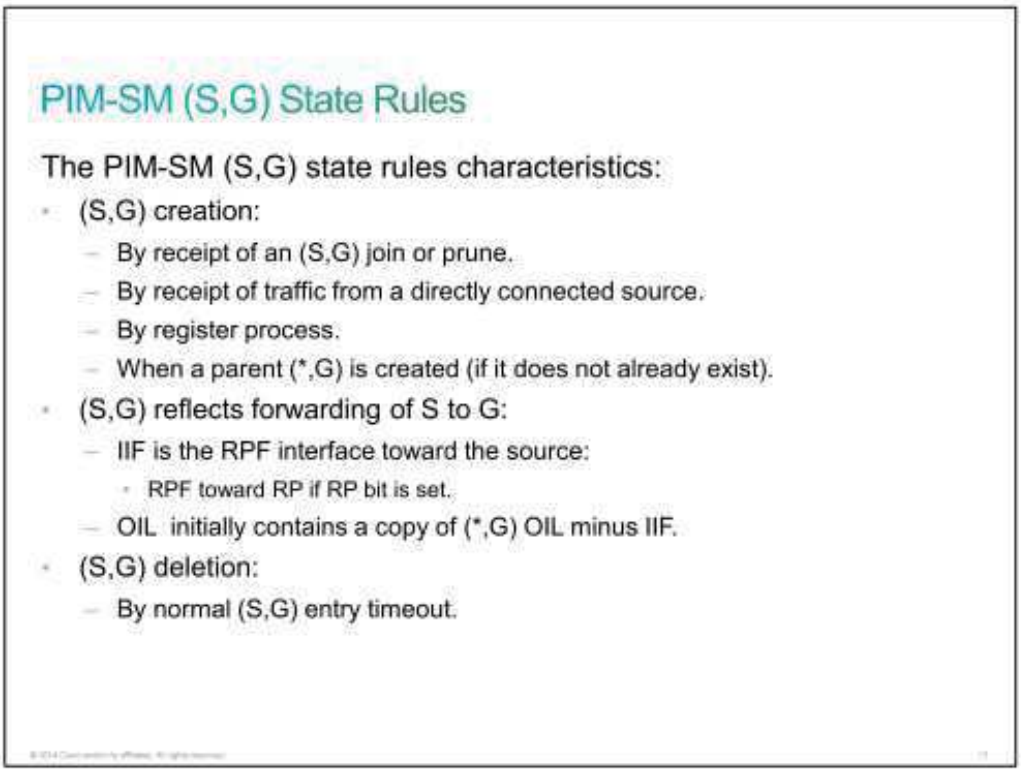


https://t.me/learningnets

- (*,G) entries are deleted when the expire timer counts down to 0. This action only occurs when one of these conditions exists:
  - The OIL is null.
  - No child (S,G) entry exists.

## PIM-SM (S,G) State Rules

The PIM-SM (S,G) state rules are described below:




In PIM-SM, the (S,G) state is created as a result of one of the following:

- The receipt of an (S,G) join or prune message.
- The receipt by the first-hop router of a packet from a directly connected source. This action will trigger the PIM-SM register process.
- The following events occur for (S,G) entry creation:
  - If a corresponding (*,G) entry does not exist, the (*,G) must be created first before the (S,G) entry can be created.
  - The RPF information is computed for the source S. This information is stored in the (S,G) entry as the IIF and the RPF neighbor (that is, the PIM neighbor in the direction of the source).

The exception to this rule occurs when the RP bit (R flag) is set in the (S,G) entry and the RPF interface is pointing up the shared tree. This occurs after the SPT switchover and the router receives the RP-bit prune message to prune the (S,G) traffic from the shared tree. This mechanism allows duplicate (S,G) traffic to be blocked from flowing down the shared tree after a downstream router has switched to the SPT.

The OIL of the (S,G) entry is populated with a copy of the OIL from the parent (*,G) entry, without the IIF. (The IIF must not appear in the OIL, or a multicast route loop may occur.)



https://t.me/learningnets

In PIM-SM, the (S,G) entries are deleted when their expire timer counts down to 0. The expire timer is reset whenever an (S,G) packet is received and forwarded.

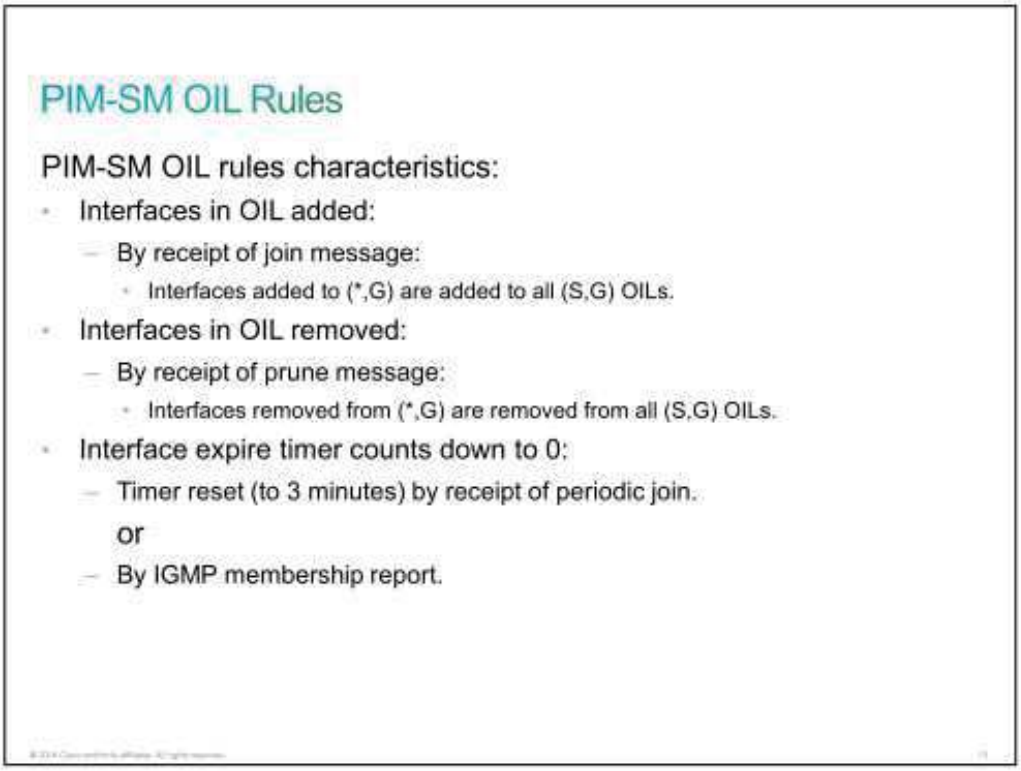
Flags behave differently based on the particular Cisco IOS, IOS XE, or IOS XR Software release.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# PIM-SM OIL Rules

The PIM-SM OIL rules are described below:




The PIM-SM OIL rules are as follows:

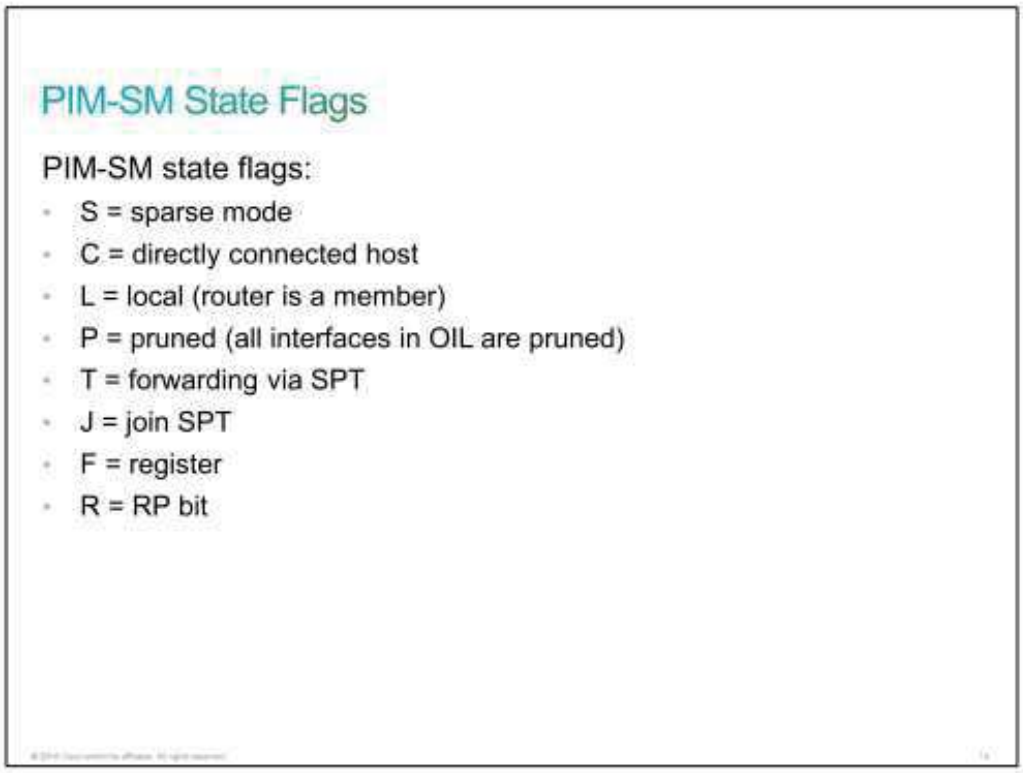
- **Adding an interface:** A check is always made to prevent the IIF from appearing in the OIL:
  - Interfaces are added to the (S,G) OIL when an (S,G) join message is received on an interface.
  - Interfaces are added to the (*,G) OIL when a (*,G) join message is received on an interface.
  - Anytime an interface is added to the (*,G) OIL, the interface is added to the OIL of all associated (S,G) OILs.
- **Removing an interface:** Interfaces removed from the (*,G) OIL are also removed from the (S,G) OILs.
  - Interfaces are removed from the OIL of a (*,G) or (S,G) entry if the interface expire timer counts down to 0. The interface expire timer is reset to 3 minutes by the receipt of periodic join messages or by an IGMP report. The downstream router will send one join message per minute. Directly connected members on the interface will send an IGMP report.
  - Interfaces are removed from the OIL if a prune message is received and if another router on the same interface (if the interface is connected to a multiaccess network like Ethernet) does not override the message.



https://t.me/learningnets

# PIM-SM State Flags

The PIM-SM state flags are described below:




The most common PIM-SM state flags are as follows:

- **S flag:** Indicates that the group is operating in sparse mode (appears only on [*,G] entries).
- **C flag:** Indicates that there is a member of the group that is directly connected to the router.
- **L flag:** Indicates that the router itself is a member of this group and is receiving the traffic. This condition would exist for the Auto-RP discovery group 224.0.1.40, which all Cisco routers join automatically.
- **P flag:** Set whenever all interfaces in the OIL of an entry are pruned or the list is null. This action generally means that the router will send prune messages to the RPF neighbor to try to shut off this traffic.
- **T flag:** Indicates that at least one packet was received and forwarded via the SPT. This flag only appears on (S,G) entries.
- **J flag (join SPT):** When this flag is set in the (*,G) entry, it indicates that the rate of traffic flowing down the shared tree is above the SPT threshold. This state will cause a switch to the SPT for the next packet that is received down the shared tree. When this flag is set in the (S,G) entry, it indicates that the (S,G) entry—and hence the SPT—was created because the SPT threshold was exceeded. If the rate of the (S,G) traffic drops back below the SPT, the router will attempt to switch this traffic flow back to the shared tree.
- **F flag (register):** This flag is set on an (S,G) entry when the source is directly connected to the router. This setting indicates that this router is a first-hop router, and it triggers the router to send register messages to the RP to inform the RP of this active source. The F flag is also set on the (*,G) entry if any associated (S,G) entries have the F flag set.



https://t.me/learningnets

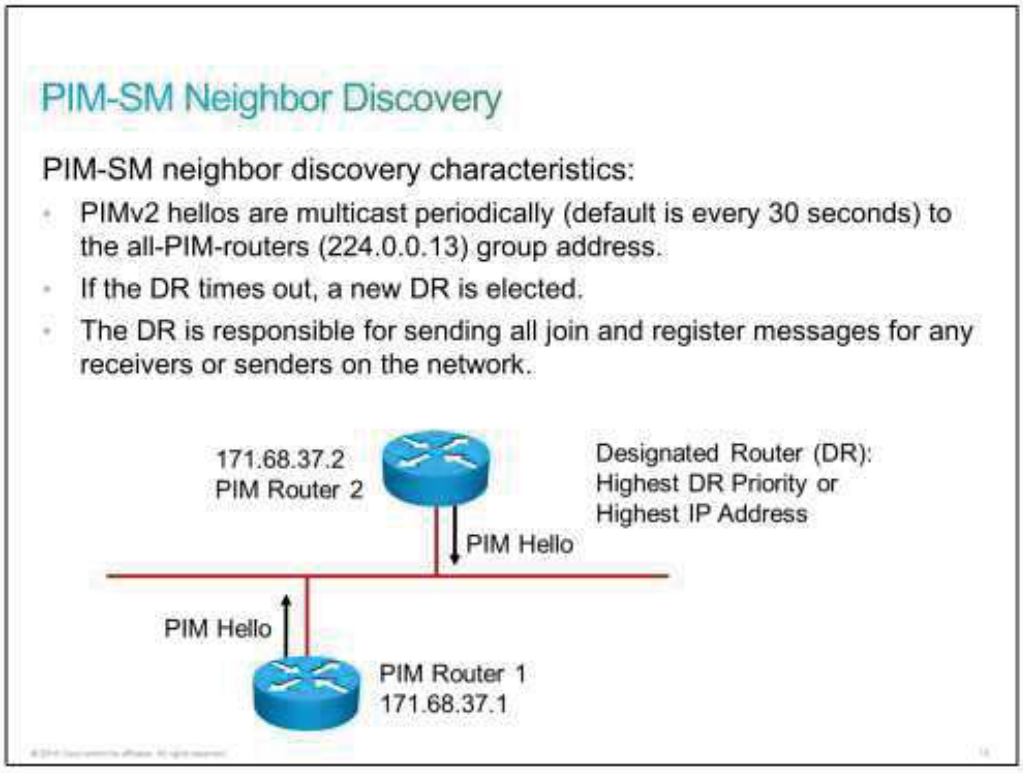
- **R flag (RP bit):** This flag is set on (S,G) entries only and indicates that the (S,G) forwarding information in the entry is applicable to (S,G) traffic flowing down the shared tree. The R flag is set on an (S,G) entry by the receipt of an (S,G) RP-bit prune message. Downstream routers on the shared tree will send messages to request that specific (S,G) traffic flow be pruned off the shared tree. This action is done to eliminate duplicate (S,G) traffic after a downstream router has switched to the SPT. Whenever the R flag is set on an (S,G) entry, the RPF information must be changed to point toward the RP instead of pointing toward source S. This action is taken because the (S,G) entry is now applicable to (S,G) traffic arriving down the shared tree. As a result, the RPF information must point up the shared tree in order for arriving (S,G) packets to pass the RPF check.



https://t.me/learningnets

# PIM-SM Neighbor Discovery

PIM hello messages are sent periodically to discover the existence of other PIM routers on the network. They are also used to elect the DR on multiaccess networks.




On multiaccess networks (for example, Gigabit Ethernet), the PIM hello messages are sent via multicast to the all-PIM-routers (224.0.0.13) multicast group address.

If there is more than one router on a multiaccess network, a DR is elected. In PIM-SM networks, the DR is responsible for sending joins to the RP for members on the multiaccess network. It is also responsible for sending registers to the RP for sources on the multiaccess network.

To elect the DR, each PIM node on a multiaccess network examines the received PIM hello messages from its neighbors. The DR will first be selected based on the highest DR priority (the default is 1). If the PIM DR priority is the same, the router compares the IP address of its interface with the IP address of its PIM neighbors. The PIM neighbor with the highest IP address is elected the DR.

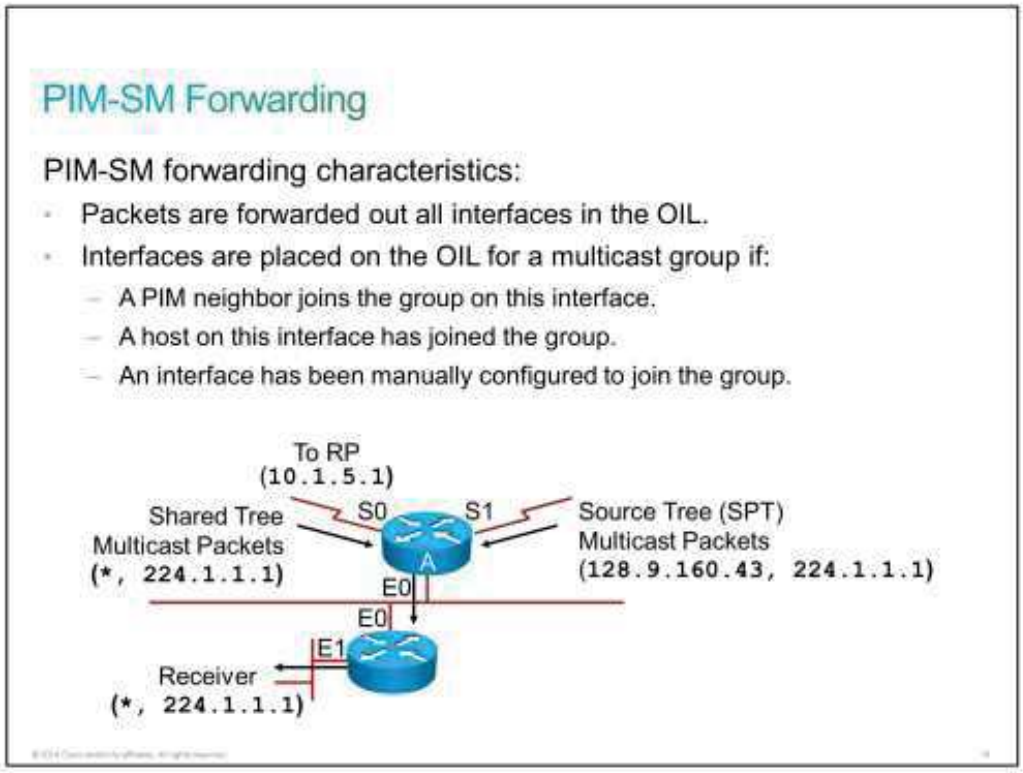
If no PIM hellos have been received from the elected DR after some period (this variable is configurable), the DR election mechanism is run again to elect a new DR.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# PIM-SM Forwarding

In PIM-SM, the incoming multicast packets for group G are forwarded to all interfaces that are in the multicast routing table entry OIL. An attempt is made to forward packets by using a matching (S,G) entry, if one exists. In that case, the OIL of the (S,G) entry is used to control the forwarding of the multicast packet. If no (S,G) entry exists, then the OIL of the (*,G) entry is used to control the forwarding of the multicast packet.




Interfaces are placed in the (*,G) OIL if and only if one of the following events has occurred:

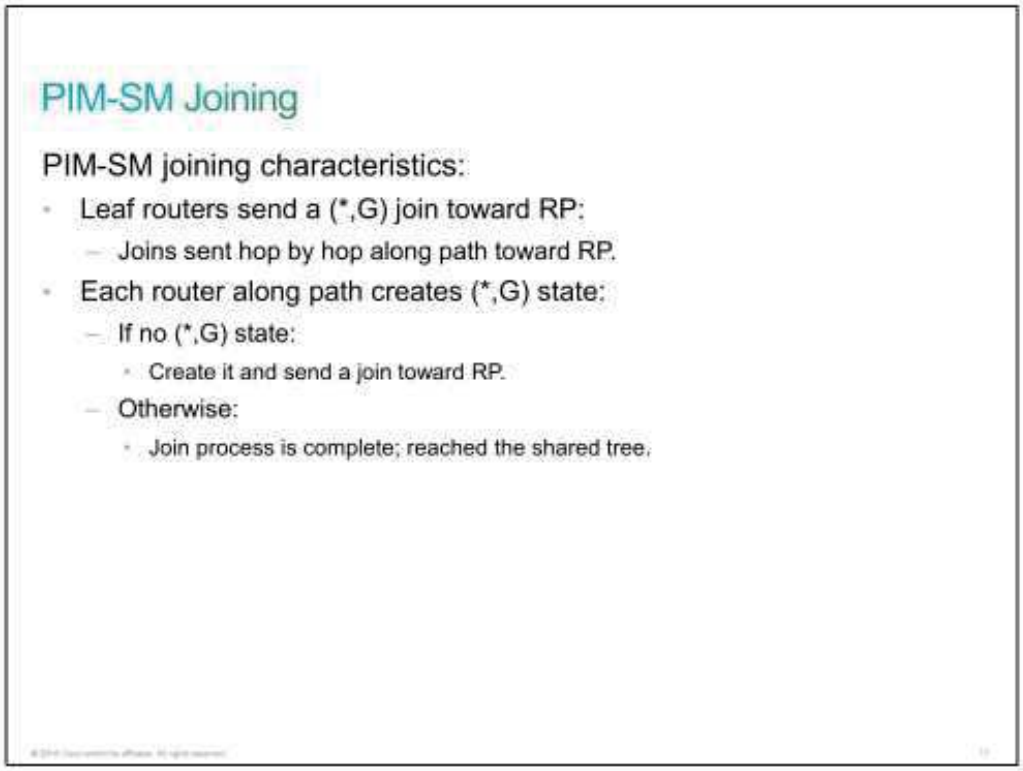
- A router has received a PIM join for this group from another PIM neighbor via this interface.
- A host on this interface has joined the group via IGMP.
- A router interface has been manually configured to join a group.



https://t.me/learningnets

# PIM-SM Joining

When a last-hop router wants to begin receiving multicast traffic for group G, it sends a PIM (*,G) join message to its upstream PIM neighbor in the direction of the RP.




Like all PIM control messages, the join is multicast, with a TTL of 1, to the multicast address (224.0.0.13 in PIMv2), which is the all-pim-routers group. The IP address of the upstream PIM neighbor that needs to perform the join is indicated in the body of the multicast PIM join message. This action permits all PIM routers on a multiaccess network to receive the join, but only the indicated upstream PIM neighbor will take action on the join.

When a PIM router receives a (*,G) join for group G from one of its downstream PIM neighbors, it will check to see whether any (*,G) state exists for group G in its multicast routing table.

PIM-SM joining characteristics:

- If a (*,G) state for group G exists, then the interface from which the join was received is placed on the (*,G) OIL.
- If no (*,G) state for group G exists, a (*,G) entry is created. Then, the interface from which the join was received is placed in the (*,G) OIL, and a (*,G) join is sent toward the RP.

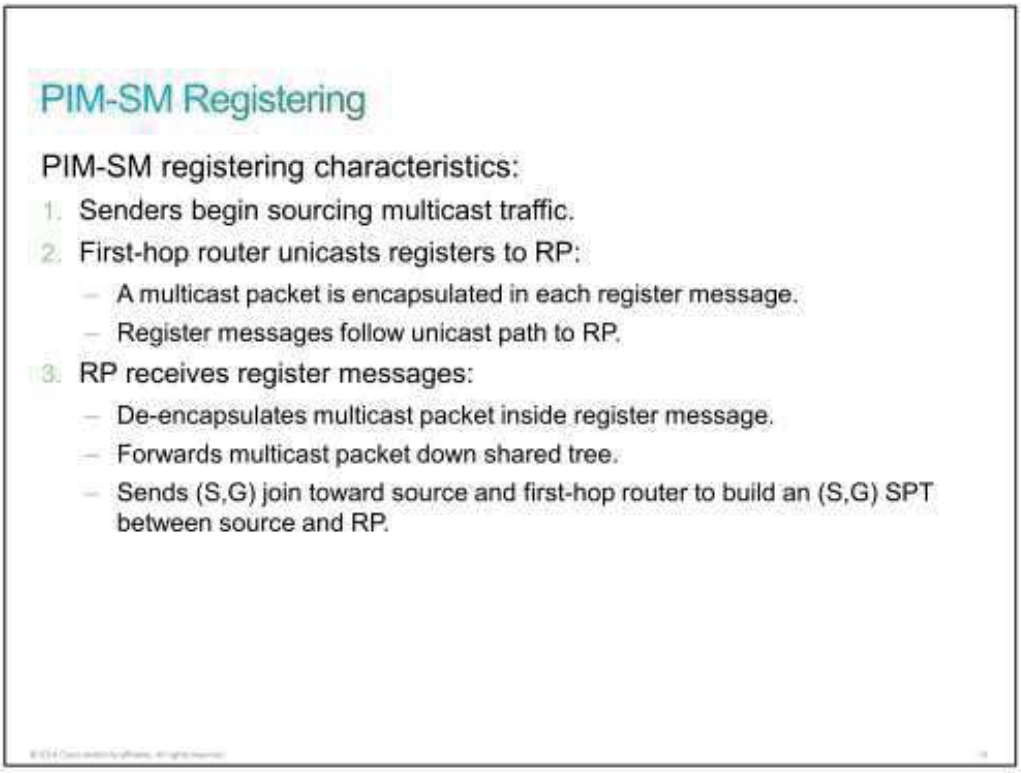
The result of the PIM join mechanism is to create a (*,G) state from the last-hop router to the RP. This action allows group G multicast traffic to flow down the shared tree to the last-hop router. Joins and prunes are sent in PIM join and prune messages, each of which may contain multiple joins and multiple prunes.



https://t.me/learningnets

# PIM-SM Registering

With a source-only host, it is permissible (and normal) for the host to simply begin sending multicast traffic without ever joining the group via IGMP. In PIM-SM, when a first-hop router receives the first multicast packet from directly connected source S for group G, it first creates an (S,G) state. It then sets the F bit in the (S,G) entry to indicate that it is a directly connected source. Additionally, it also sets the registering flag to indicate that it is in the process of registering the source to the RP.






https://t.me/learningnets

## PIM-SM Registering (Cont.)

4. First-hop router receives (S,G) join:
   - SPT between source and RP now built.
   - Begins forwarding traffic down (S,G) SPT to RP.
   - (S,G) traffic temporarily flowing down two paths to RP.
5. RP receives traffic down native (S,G) SPT:
   - Sends a register-stop message to source and first-hop router.
6. First-hop router receives register-stop:
   - Stops encapsulating traffic in register messages.
   - (S,G) traffic now flowing down single SPT to RP.

With a source-only host, it is permissible (and normal) for the host to simply begin sending multicast traffic without ever joining the group via IGMP. In PIM-SM, when a first-hop router receives the first multicast packet from directly connected source S for group G, it first creates an (S,G) state. It then sets the F bit in the (S,G) entry to indicate that it is a directly connected source. Additionally, it also sets the registering flag to indicate that it is in the process of registering the source to the RP.

The first-hop router encapsulates the original multicast packet in a PIM register message and unicasts it to the RP. Any subsequent multicast packets that are received from directly connected source S for group G are also encapsulated in a register message and unicast to the RP. The process of sending encapsulated traffic continues until an (S,G) register-stop message is received from the RP.

When the RP receives the register message, it de-encapsulates the message. If this packet is being forwarded to a group for which the RP has a (*,G) state, the RP will forward the original packet to all of the interfaces in the (*,G) entry OIL, create an (S,G) state and send an (S,G) join back toward the source to build a branch of an (S,G) SPT from source S to the RP.

When the first-hop router receives the (S,G) join (sent hop by hop from the RP), it normally processes it by adding the interface from which the join was received to the OIL of the existing (S,G) entry. This entry was originally created when the first-hop router received the first multicast packet from the directly connected source. At this time, the building of a branch of the SPT from the source to the RP is complete. Now the first-hop router can start forwarding multicast traffic to the RP via the newly built branch of the (S,G) SPT.

(S,G) traffic temporarily flows to the RP in two ways: via register messages (until a register-stop message is received) and via the native SPT. When the RP begins receiving (S,G) traffic natively (that is, traffic that is not encapsulated in register messages) down the SPT, the RP will set the T bit in the (S,G) entry. This activity denotes that traffic is successfully flowing down the SPT.

When any (S,G) register messages are received by the RP, the RP recognizes that the T bit is set in the (S,G) entry. It will now respond by sending a PIM (S,G) register-stop message to the first-hop router. This action notifies the first-hop router that traffic is now being received natively down the SPT.

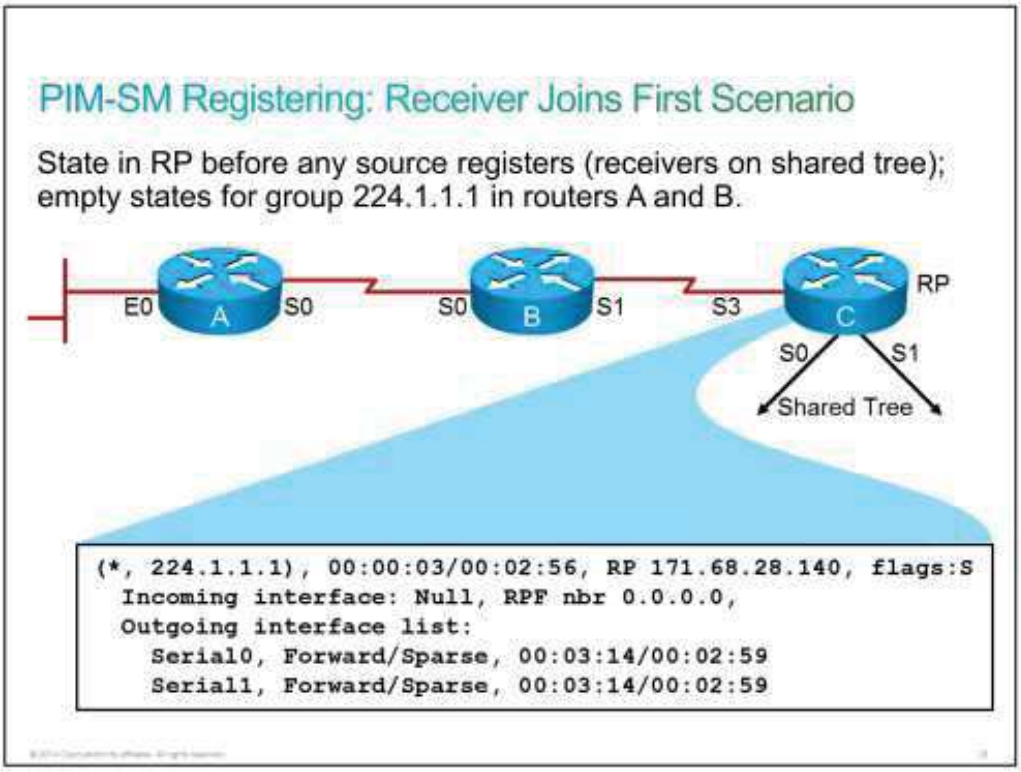
When the (S,G) register-stop message is received by the first-hop router, it clears the registering flag in the (S,G) entry and stops encapsulating (S,G) traffic in register messages. Traffic is now only flowing down the SPT to the RP.



https://t.me/learningnets

# PIM-SM Registering: Receiver Joins First Scenario

The first scenario of the register process pertains to a situation where the shared tree exists, meaning that the receivers joined the group before the source started sending the multicast traffic.




This figure shows the state in the RP before the registration of the sender in PIM-SM. Note that some receivers have already joined the shared tree—which is why the RP router already has the (*,G) entry:

- The incoming interface is null and the RPF neighbor is 0.0.0.0. This data indicates that router C is the RP.
- The OIL contains S0 and S1, which are assumed to be the only two active branches of the shared tree.

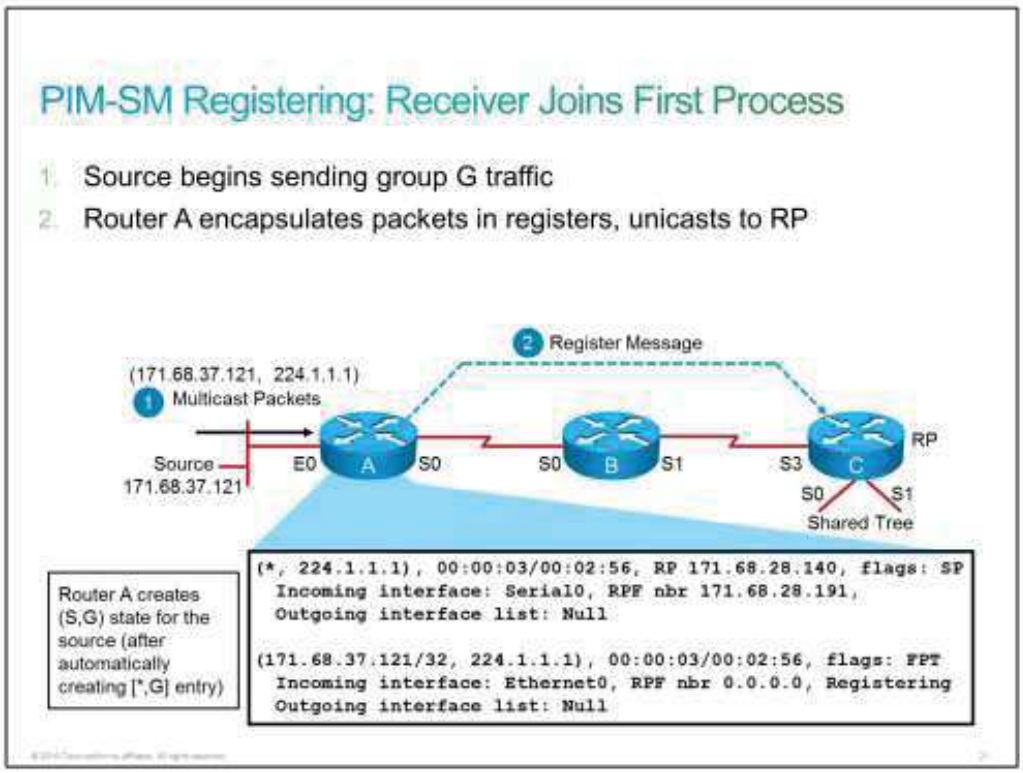
Currently, in routers A and B, there is no state for group 224.1.1.1.

## PIM-SM Registering: Receiver Joins First Process

When the first packet arrives from the source, the first-hop router A creates an (S,G) state. A (*,G) entry must be created before the (S,G) entry may be created.



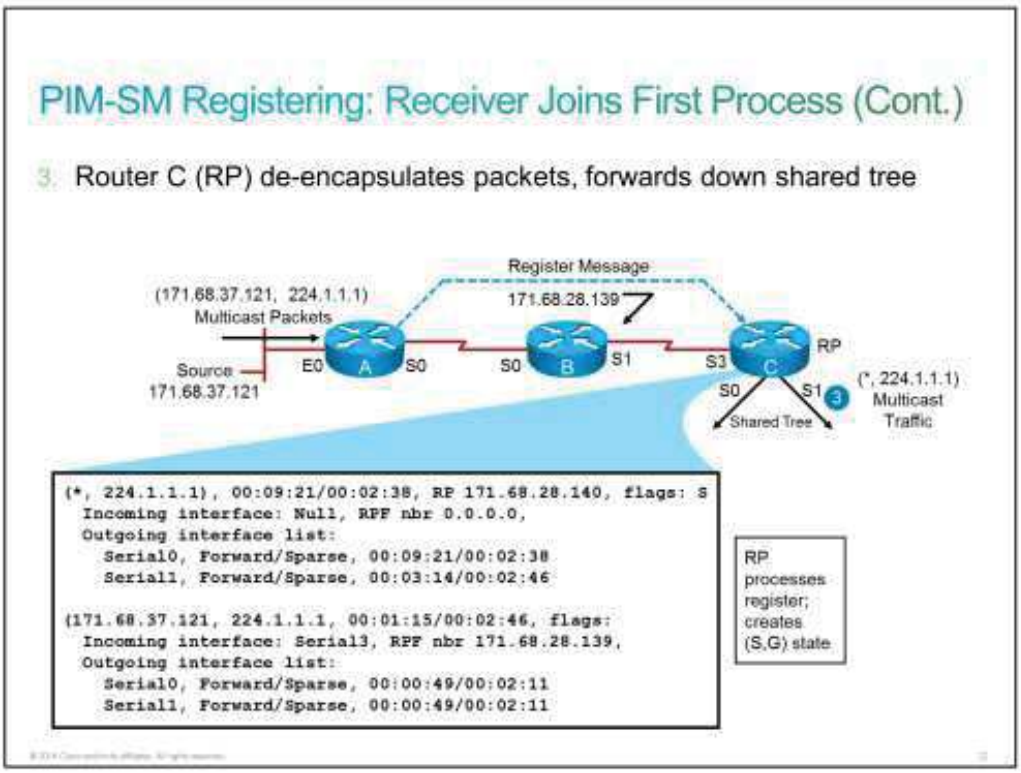
https://t.me/learningnets




1. Observe the following from router A show output shown in the figure:
   - RPF information for the (*,224.1.1.1) entry points up the shared tree via S0 with the RPF neighbor of 171.68.28.191 (S0 of router B).
   - In this figure, no members have joined the (*,G) group, therefore, the OIL of the (*,G) entry is null.
   - Because the OIL is null, the P flag (pruned) is set.
   - The (171.68.37.121, 224.1.1.1) entry is then created. Note the following:
     - RPF information for this entry points toward the source via E0. The RPF neighbor is 0.0.0.0 because the source is directly connected.
     - The F register flag is set in the (S,G) entry, which indicates that this is a directly connected source.
     - The (S,G) OIL receives a copy of the (*,G) OIL, which is null.
     - The registering flag is set in the (S,G) entry, which indicates that the router is sending register messages to the RP for this source.
     - Because the OIL is null, the P flag (pruned) is set.
2. When the (*,G) and (S,G) entries have been created, the first-hop router encapsulates the multicast packets in PIM register messages, and unicasts them to the RP.



https://t.me/learningnets



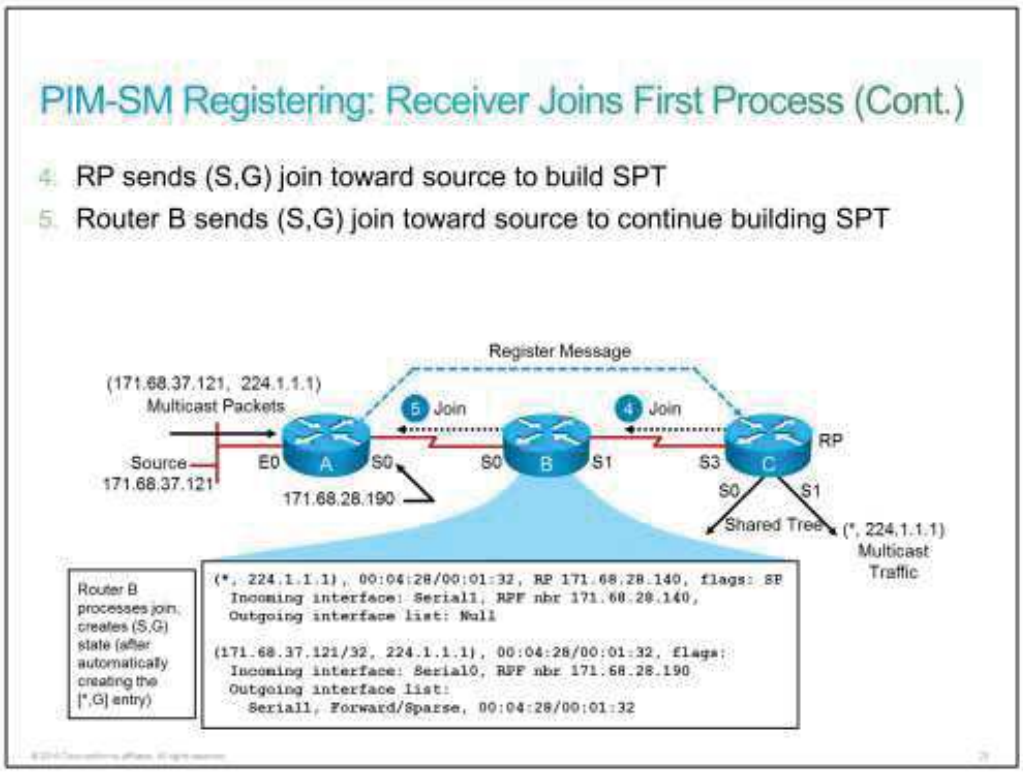

3. When the first encapsulated multicast packet arrives in a register message to the RP, the RP creates an (S,G) state as follows:
   - RPF information is calculated using the source address that is contained in the multicast packet that is encapsulated inside of the register message. This action results in an IIF of S3 and an RPF neighbor of 171.68.28.139 (Router B S1 interface).
   - Next, the OIL of the parent (*,G) entry is copied into the OIL of the new (S,G) entry. An additional check is made to ensure that the IIF (S3) does not appear in the OIL. If it does, it is removed to prevent a route loop.

The router is now ready to forward the (S,G) packet that was encapsulated in the register message using the newly created (S,G) state. Traffic is always forwarded using the matching (S,G) entry, if one exists. Otherwise, the (*,G) entry is used. This action is accomplished as follows:

- Because this packet was received inside of a register message, the RPF check is bypassed.
- Next, the router forwards a copy of the packet to all interfaces in the (S,G) OIL. In this case, a copy is sent to S0 and S1, which correspond to the two branches of the shared tree.
- The T flag is not yet set in the (S,G) entry. However, when the first (S,G) packet is received natively (via the IIF) and forwarded using this entry, the T flag will be set.



https://t.me/learningnets

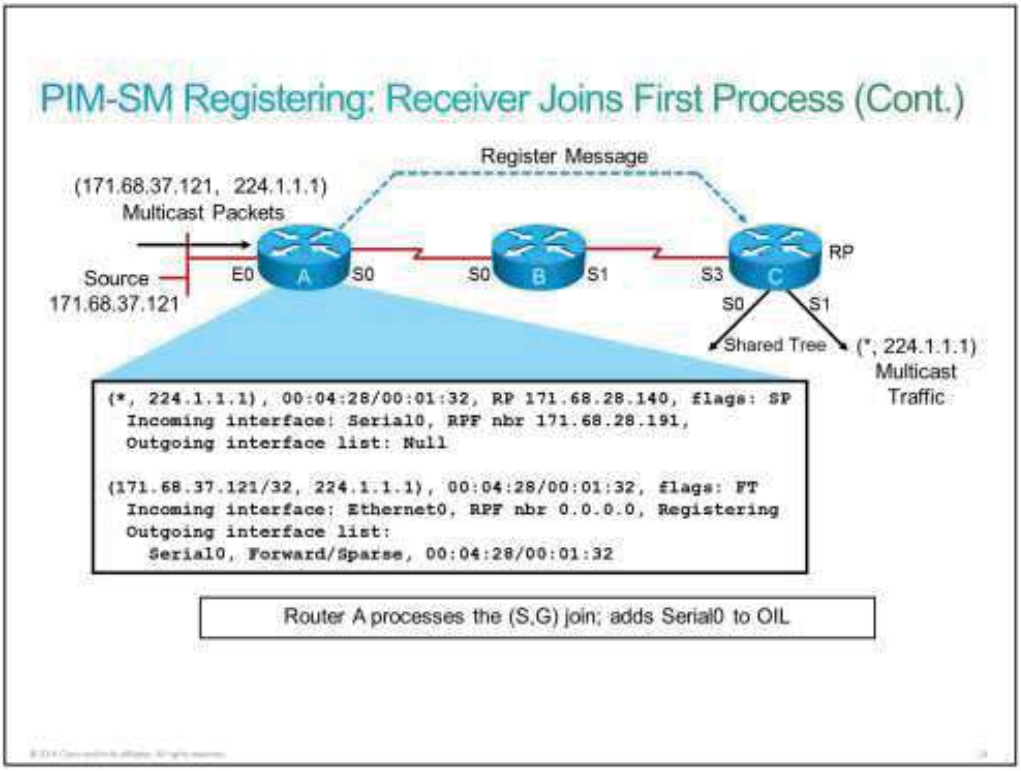



4. Because the RP has an existing (*,G) state, it sends an (S,G) join toward source S to build a branch of the (S,G) SPT from the source to the RP. The IP address of S was found in the IP header of the encapsulated multicast packet.
5. The next upstream router in the direction of the source (router B) processes the received (S,G) join and creates an (S,G) state. Router B must create the (*,G) state before the (S,G) entry may be created. Note the following:
   - RPF information for the (*,G) entry for router B points up the shared tree via S1 with the RPF neighbor of 171.68.28.140 (S3 of the RP.)
   - In this figure, no members have joined the group; therefore the OIL of the (*,G) entry is null.
   - Because the OIL is null, the P flag (pruned) is set.
   - The (S,G) entry is then created. Note the following:
     - RPF information for this entry points toward the source via S0. The RPF neighbor is 171.68.28.190 (S0 of router A).
     - The (S,G) OIL initially receives a copy of the (*,G) OIL, which is null.
     - Interface S1, which is the interface that received the (S,G) join, is added to the (S,G) OIL.
     - The T flag is not yet set in the (S,G) entry. However, when the first (S,G) packet is forwarded using this entry, the T flag will be set.

Because the OIL of the (S,G) transitioned from null to non-null (when router B added S1 to the OIL of the newly created entry), a PIM (S,G) join is sent to router A to continue the process of joining the SPT.

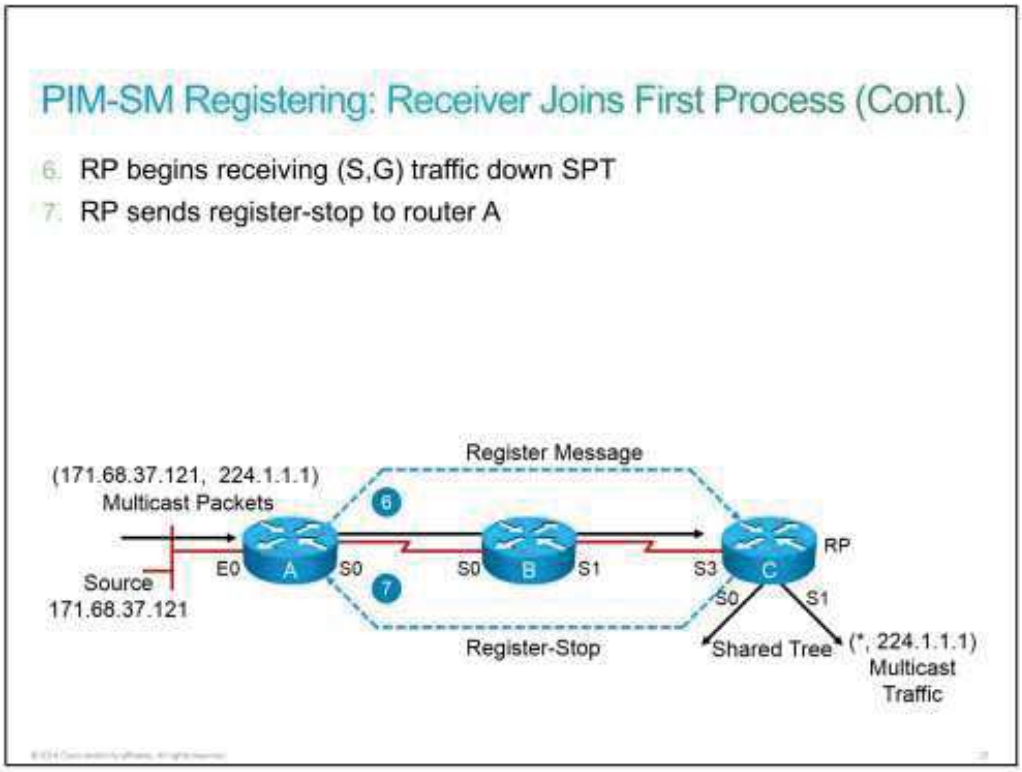


https://t.me/learningnets




When the first-hop router A receives the (S,G) join, it checks its state. Because an (S,G) entry exists, router A simply adds the interface on which the (S,G) join was received to the OIL. This action results in the following:

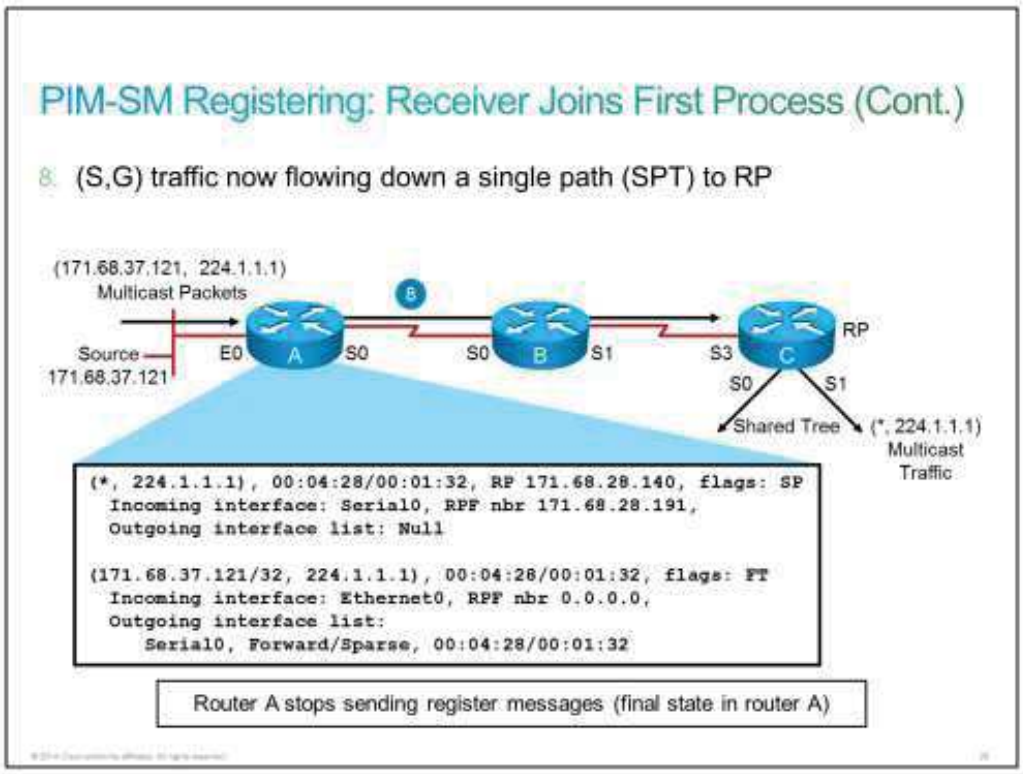
- S0 is now listed in the OIL, because the RP joined the SPT via this interface.
- The P flag is cleared, because the OIL is no longer null.






https://t.me/learningnets

6. At this point, a branch of the (S,G) SPT has been built between the first-hop router and the RP, and (S,G) traffic begins flowing down the SPT. The RP is now receiving the (S,G) traffic natively down this branch of the SPT. This activity causes the T flags to set in the (S,G) entries along this path, including in the RP. The branch of the SPT—in fact, all PIM tree branches—are built in the opposite direction by propagating joins from the leaves of the tree to the root of the tree. In this case, (S,G) joins were sent from the RP upstream toward the first-hop router that is connected to the source.
7. Because the T flag was set in the (S,G) entry, the next register message that is received by the RP will cause the RP to send an (S,G) register-stop message to the first-hop router. The register-stop message informs the first-hop router that the encapsulation of (S,G) packets in register messages is no longer necessary.

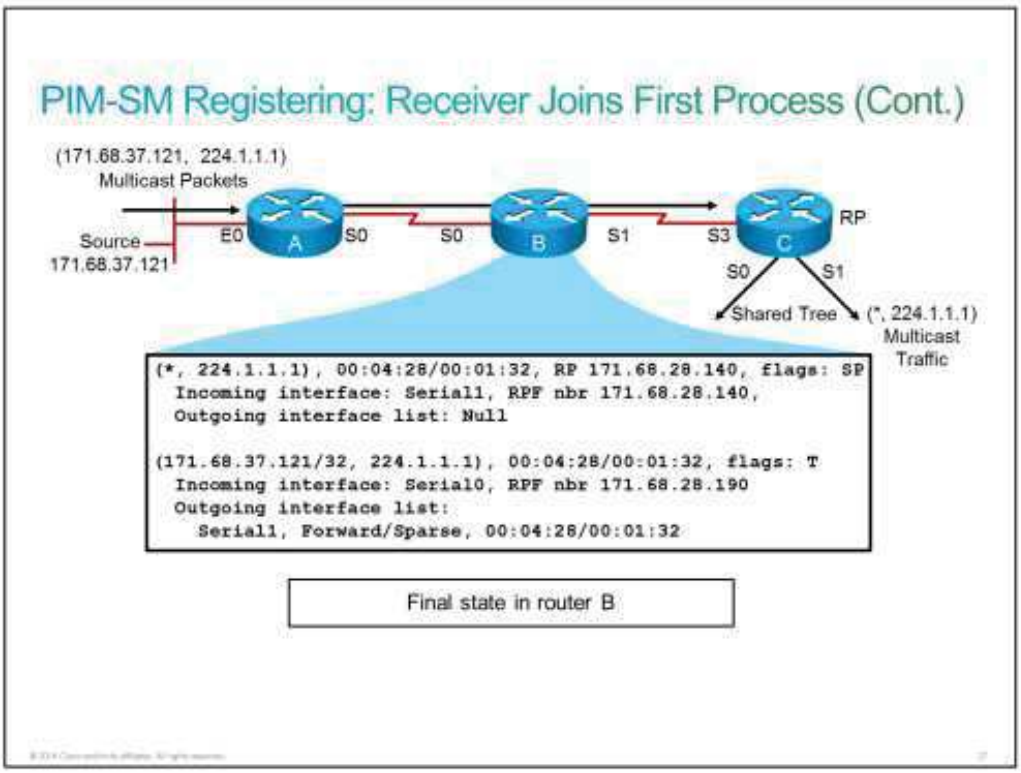



When the first-hop router A receives the (S,G) register-stop message, it ceases sending encapsulated register messages for (S,G) traffic. The word "Registering" is no longer being displayed on the second line of the (S,G) entry, indicating that router A is not sending registers. This state is the final state in router A after the registration process.

8. (S,G) traffic is now only flowing down a branch of the (S,G) SPT to the RP.

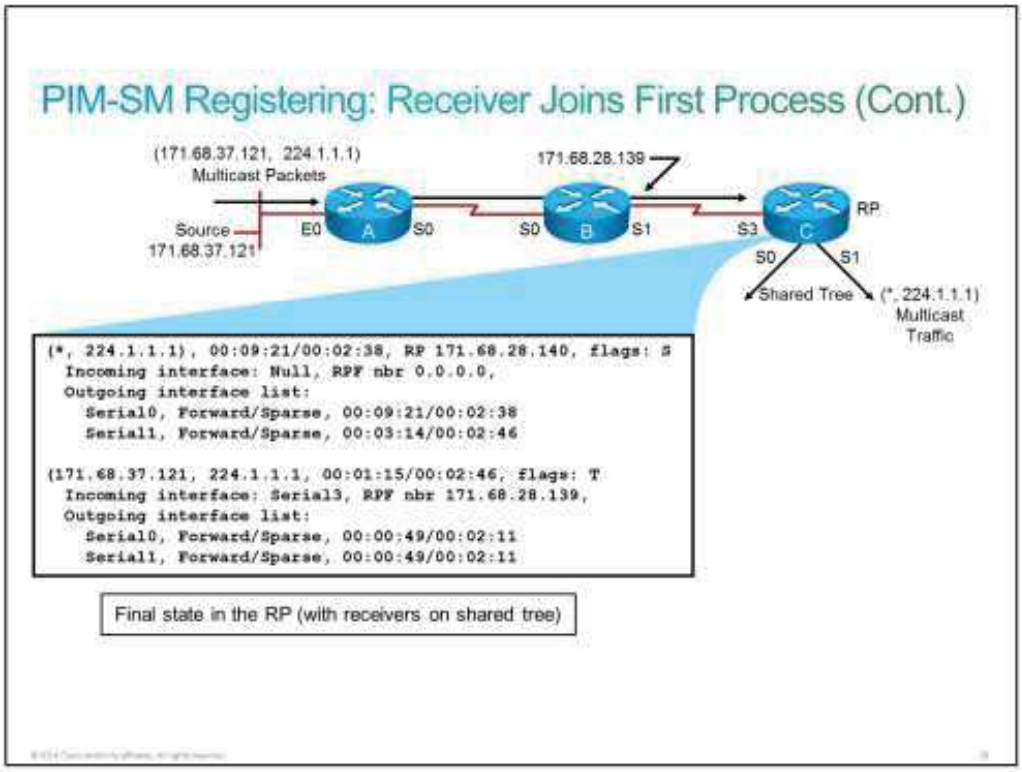


https://t.me/learningnets




When the registration process is completed, router B has the final state as shown in the figure. Note the following items in the (S,G) entry:

- The T flag is now set, indicating that (S,G) traffic is flowing along this path.
- The (*,G) entry still has a null OIL, and the P flag is still set. This state exists because there are no members that have joined the shared tree on this branch of the tree.






https://t.me/learningnets

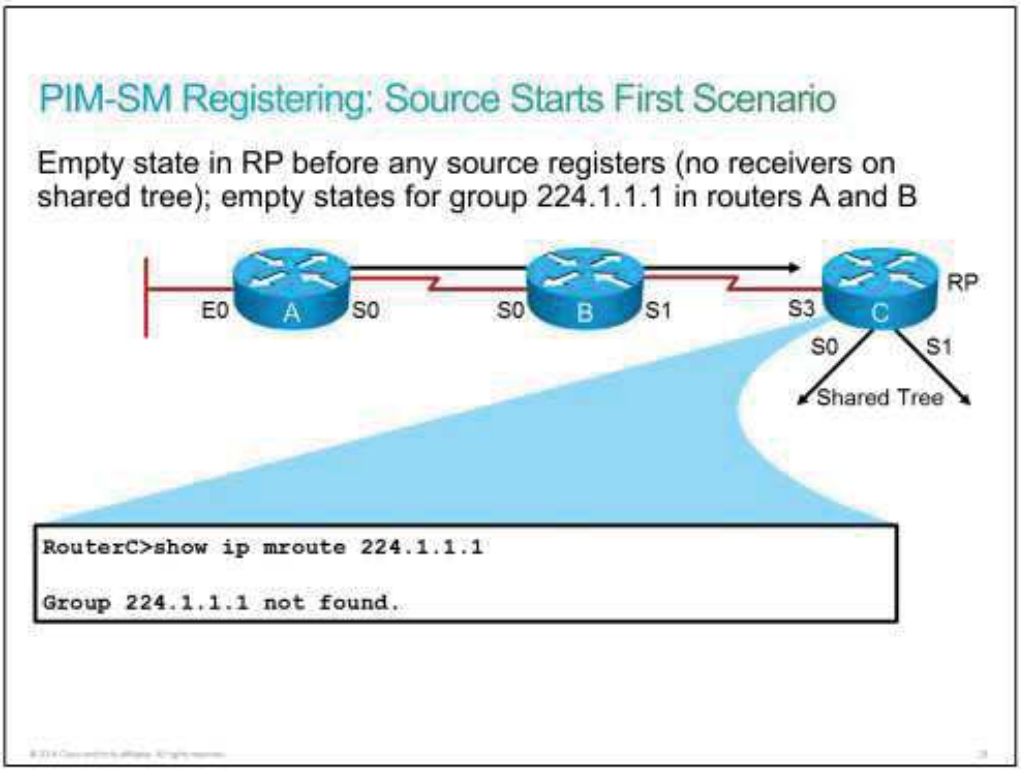
When the registration process is completed, the final state in the RP is as shown in the figure. The newly created (S,G) entry now has the T flag, indicating that (S,G) traffic is flowing along the SPT.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# PIM-SM Registering: Source Starts First Scenario

The next scenario of the register process pertains to the situation where the source starts but there are no receivers for it.



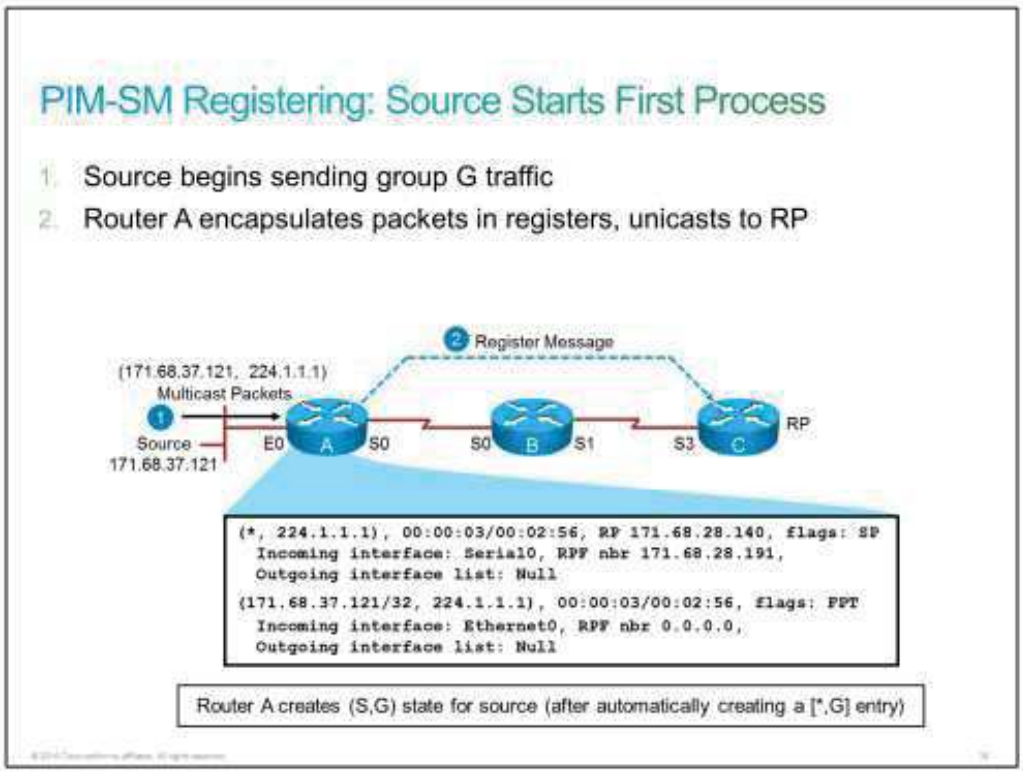

That is, no branches of a shared tree exist rooted at the respective RP. Depending on whether there are any existing receivers for group G on the shared tree (RP tree), the RP manages the register process differently. Notice that no state for group G exists in the RP, because there are no receivers on the shared tree yet. The same applies to all the other routers.

## PIM-SM Registering: Source Starts First Process

When source S begins sending traffic to group G, the first-hop router A creates a (*,G) and an (S,G) state.



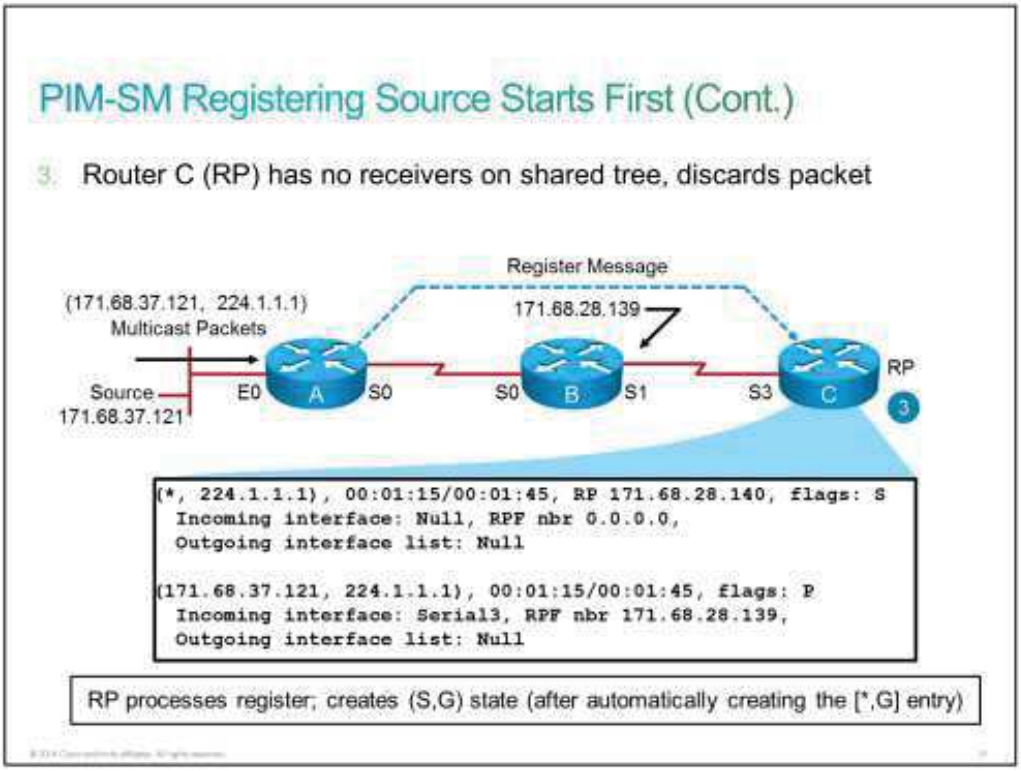
https://t.me/learningnets




1. A (*,G) entry must be created before the (S,G) entry may be created. Note the following:
   - RPF information for this entry points up the shared tree via S0 with the RPF neighbor of 171.68.28.191 (S0 of router B).
   - In this figure, no members have joined the (*,G) group, therefore the OIL of the (*,G) entry is null.
   - Because the OIL is null, the P flag (pruned) is set.
   - The (S,G) entry is then created. Note the following:
     - RPF information for this entry points toward the source via E0. The RPF neighbor is 0.0.0.0 because the source is directly connected.
     - The (S,G) OIL receives a copy of the (*,G) OIL, which is null.
     - The F flag is set in the (S,G) entry to indicate that this is a directly connected source.
     - The registering flag is set in the (S,G) entry, which indicates that the first-hop router is still sending register messages to the RP for this source.
     - Because the OIL is null, the P flag (pruned) is set.
2. After creating the entries, the first-hop router encapsulates the multicast packets in PIM register messages and unicasts them to the RP. The RP (router C) de-encapsulates the (S,G) packet and creates the (*,G) and (S,G) states. Because no receivers have joined the shared tree yet, the OILs of these entries will be null. Because the OIL of the (S,G) entry (which was just created) is null, the packet is discarded.

 © 2014 Cisco Systems, Inc.

https://t.me/learningnets




When the register message is received from router A, the RP must create an (S,G) state. However, because no previous (*,G) state existed, it must be created before the (S,G) entry is created. Notice that the (*,G) OIL is null. This condition is present because the RP has not yet received any (*,G) joins for this group.

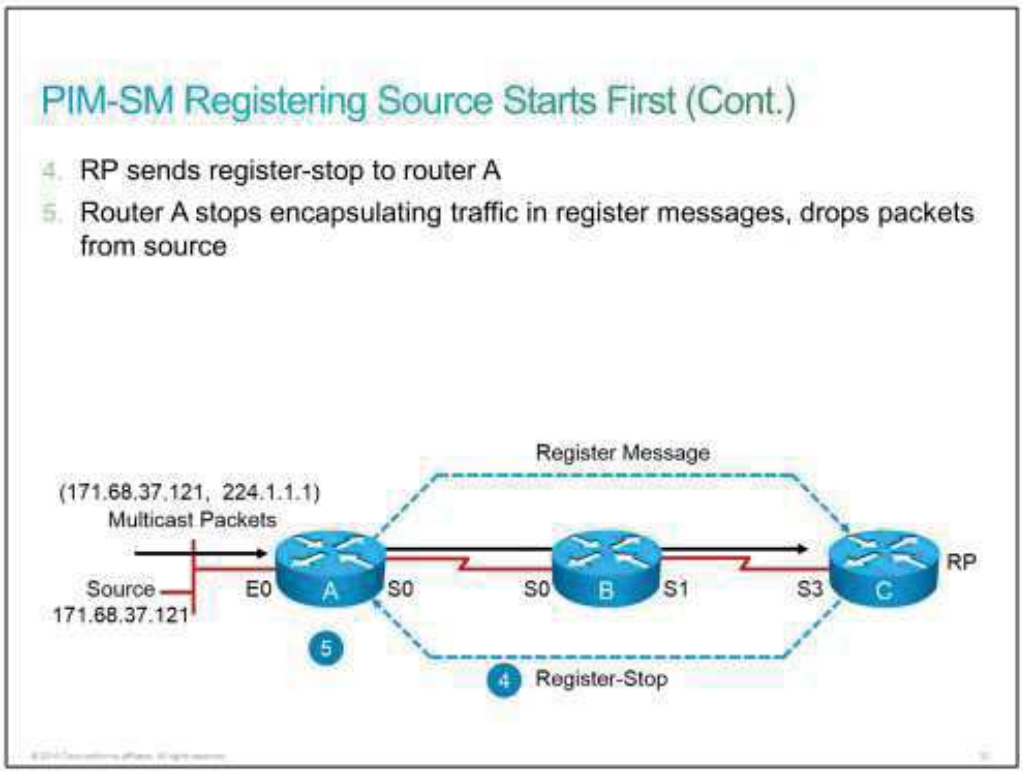
When the parent (*,G) entry has been created, the (S,G) entry may be created in the following way:

- RPF information is calculated using the source address that is contained in the multicast packet that is encapsulated inside of the register message. This action results in an IIF of S3 and an RPF neighbor of 171.68.28.139 (Router B S1 interface).
- Next, the OIL of the parent (*,G) entry is copied into the OIL of the new (S,G) entry. Because the OIL of the (*,G) entry is null, the OIL of the (S,G) entry will be null.

3. Now, the router is ready to forward the (S,G) packet that was encapsulated in the register message using the newly created (S,G) state. This activity is accomplished because the packet was received inside of a register message, so the RPF check is bypassed. The router forwards a copy of the packet to all interfaces in the matching (S,G) OIL. However, because the (S,G) OIL is null, the packet is simply discarded. The (S,G) and the parent (*,G) states remain in the RP while the source is still actively sending. After the group timer expires, the (S,G) entry is re-created. This occurs because the first-hop router will continue sending periodic register messages to the RP while the first-hop router is receiving traffic from the source.



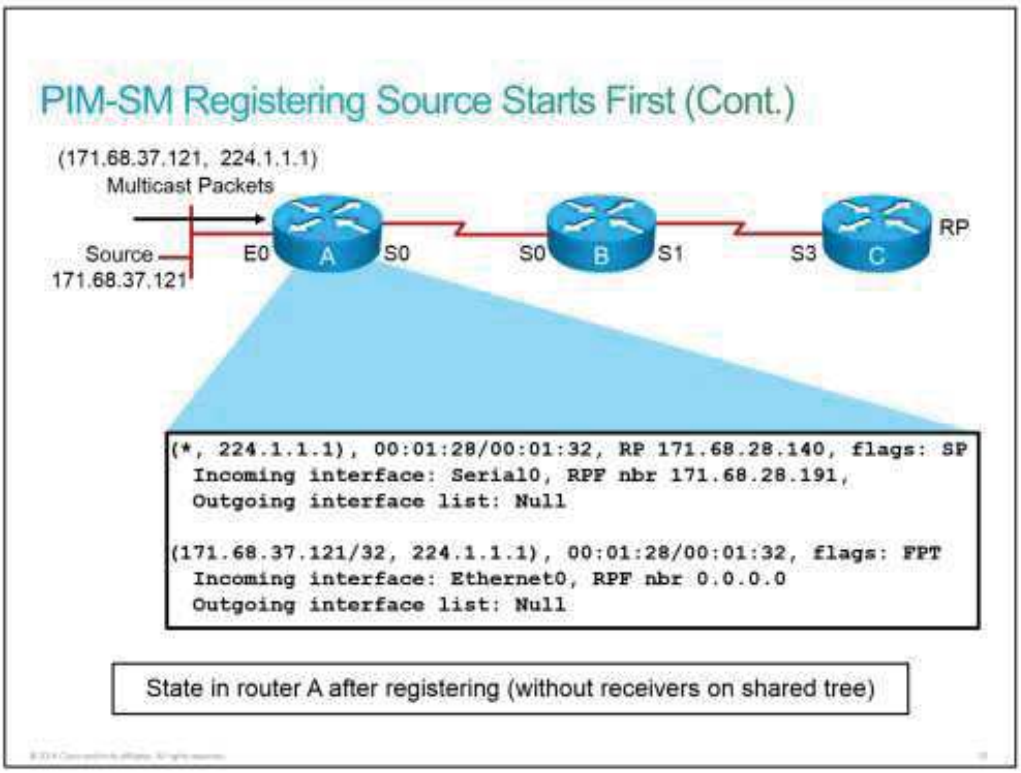
https://t.me/learningnets




4. Because the RP has no (*,G) state and, hence, no receivers on the shared tree, it does not need the (S,G) traffic. Therefore, the RP sends an (S,G) register-stop message to the first-hop router to stop sending register messages.
5. The first-hop router receives the (S,G) register-stop message and stops sending register messages for (S,G) traffic. Eventually, the original (S,G) entry will time out (after approximately 3 minutes) and be deleted. The register process will restart when the first-hop router receives the next multicast packet from the directly connected source. The RP will again respond with a register-stop, which will prevent the (S,G) traffic from flowing to the RP until it is needed.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets




By inspecting the multicast state in the first-hop router, which is created after the register-stop is received from the RP (without any receivers on a shared tree), the following may be observed in the (S,G) entry:

- The registering flag is now cleared.
- The OIL is still null, because the RP did not join the SPT.
- Because the OIL is still null, the P flag (pruned) is still set.
- Group expiration time value will count down to 0, at which time the (S,G) entry will be deleted.

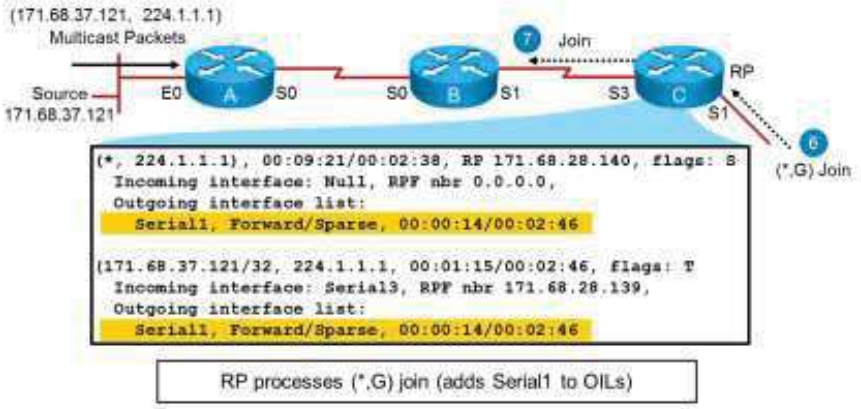
After the (S,G) entry expires, the register process restarts when the next multicast packet is received from the source. Because there were no (S,G) joins (or prunes) sent from the RP, no state currently exists in the routers between the first-hop router and the RP.



https://t.me/learningnets

## PIM-SM Registering Source Starts First (Cont.)

6. RP (router C) receives (*,G) join from a receiver on shared tree.
7. RP sends (S,G) joins for all known sources in group.




The existence of the active source S sending to group G must be maintained in the RP in the form of an (S,G) entry. This action is necessary so that the flow of (S,G) traffic may be restarted if receivers later join the shared tree for group G. This scenario is depicted in the figure.

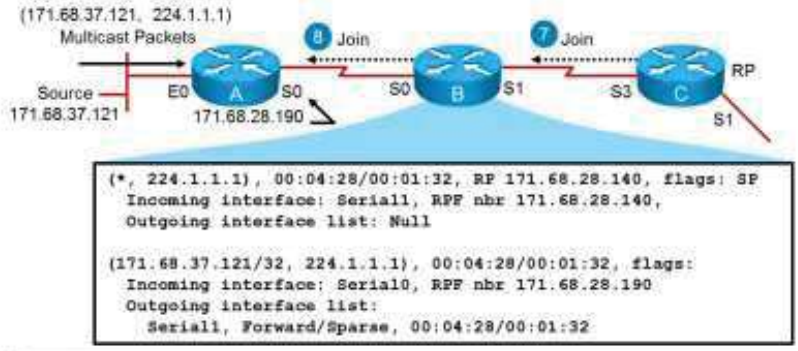
6. The RP now begins receiving (*,G) joins from last-hop routers with receivers that want to join the shared tree. The RP processes the (*,G) join in the following way:
   - S1 is added to the OIL of the (*,G) entry because a (*,G) join was received on this interface. This branch is now the only active branch of the shared tree.
   - S1 is also added to the OIL of the (S,G) entry. When the (S,G) OIL is synchronized with the parent (*,G) OIL, a check is made to ensure that the IIF of the (S,G) does not appear in the OIL of the (S,G). Such an appearance may result in a route loop.
7. The transitioning of the (*,G) OIL from null to non-null triggers the RP to scan its list of (S,G) entries for group G. The RP then sends (S,G) joins toward all sources. Each active source will build an SPT to the RP. This activity restarts the flow of (S,G) traffic to the RP.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

## PIM-SM Registering Source Starts First (Cont.)

8. Router B sends (S,G) join toward source to continue building SPT.




Router B processes join; creates (S,G) state (after automatically creating the [*,G] entry)

When router B receives the (S,G) join, it processes it by creating an (S,G) state. However, the router must first create a (*,G) state, which is accomplished as follows:

- The RPF information for the (*,G) entry points up the shared tree via S1 with the RPF neighbor of 171.68.28.140 (S3 of the RP).
- In this figure, no members have joined the group, so the OIL of the (*,G) entry is null.
- Because the OIL of the (*,G) is null, the P flag (pruned) is set.
- The (S,G) entry is then created in the following way:
  - The RPF information for the (S,G) entry points toward the source via S0. The RPF neighbor is 171.68.28.190 (S0 of router A).
  - The (S,G) OIL initially receives a copy of the (*,G) OIL, which is null.
  - Interface S1, which is the interface that received the (S,G) join, is added to the (S,G) OIL.
  - The T flag is not yet set in the (S,G) entry. However, when the first (S,G) packet is forwarded using this entry, the T flag will be set.

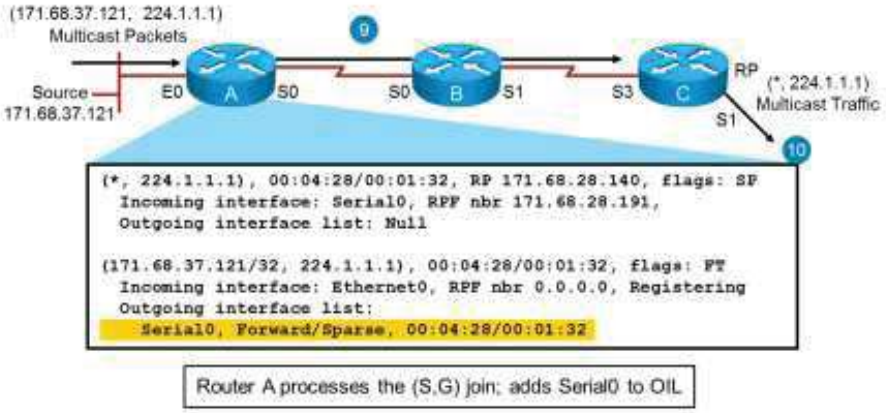
8. Because the OIL of the (S,G) transitioned from null to non-null (when router B added S1 to the OIL of the newly created entry), a PIM (S,G) join is sent to router A. This activity continues the process of joining the SPT.



https://t.me/learningnets

## PIM-SM Registering Source Starts First (Cont.)

9. RP begins receiving (S,G) traffic down SPT.
10. RP forwards (S,G) traffic down shared tree to receivers.




When the first-hop router A receives the (S,G) join, it is processed in the following way:

- S0, the interface on which the (S,G) join arrived, is added to the OIL of the existing (S,G) entry.
- The P flag (pruned) is cleared, because the OIL of the (S,G) entry is no longer null.

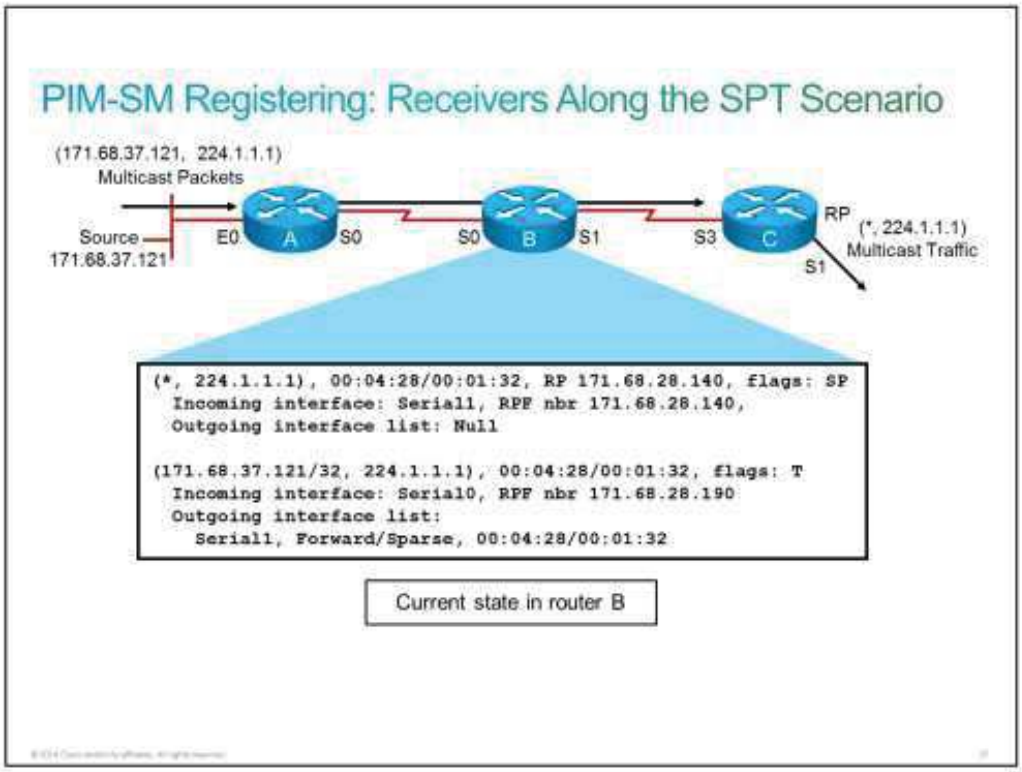
9. When S0 is added to the (S,G) OIL, traffic begins to flow down the SPT from the source to the RP.
10. The RP then forwards all incoming (S,G) traffic to the receivers down the shared tree.



https://t.me/learningnets

# PIM-SM Registering: Receivers Along the SPT Scenario

The next scenario that is presented in this section pertains to the case when there are receivers along the SPT between the first-hop router and the RP.



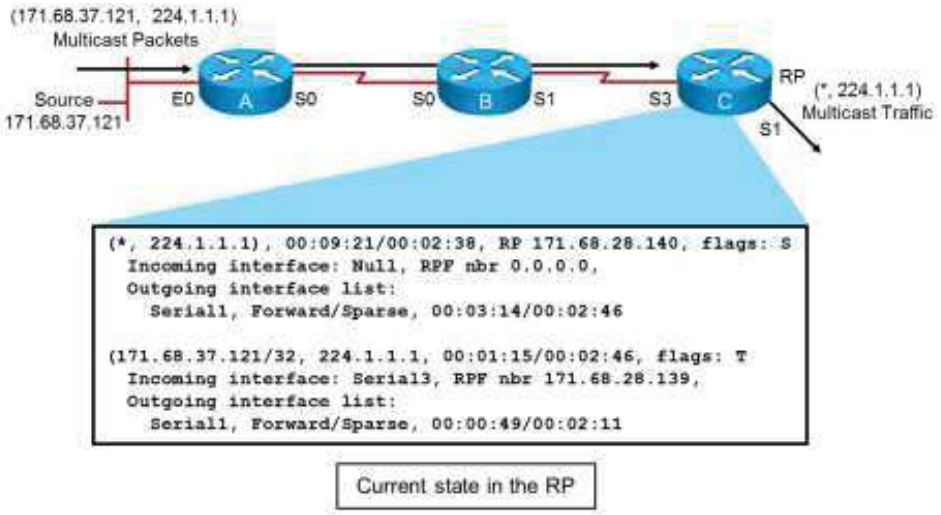

The state in router B with traffic flowing on the SPT (no receivers are yet active along this tree yet) is shown. Note the following:

- Both the (*,G) and (S,G) states were created in response to the (S,G) join being received from the RP.
- The P flag is set in the (*,G) entry, because there are no receivers on the shared tree at this point in the network.
- The T flag is set in the (S,G) entry, indicating that traffic is flowing down the tree.
- The (S,G) entry RPF nbr is the IP address of router A.
- S0 is the incoming interface of the (S,G) entry, because S0 is the RPF interface for source S via router A.
- S1 is listed in the OIL of the (S,G) entry, because the RP joined the SPT via this interface.



https://t.me/learningnets

## PIM-SM Registering: Receivers Along the SPT Scenario (Cont.)



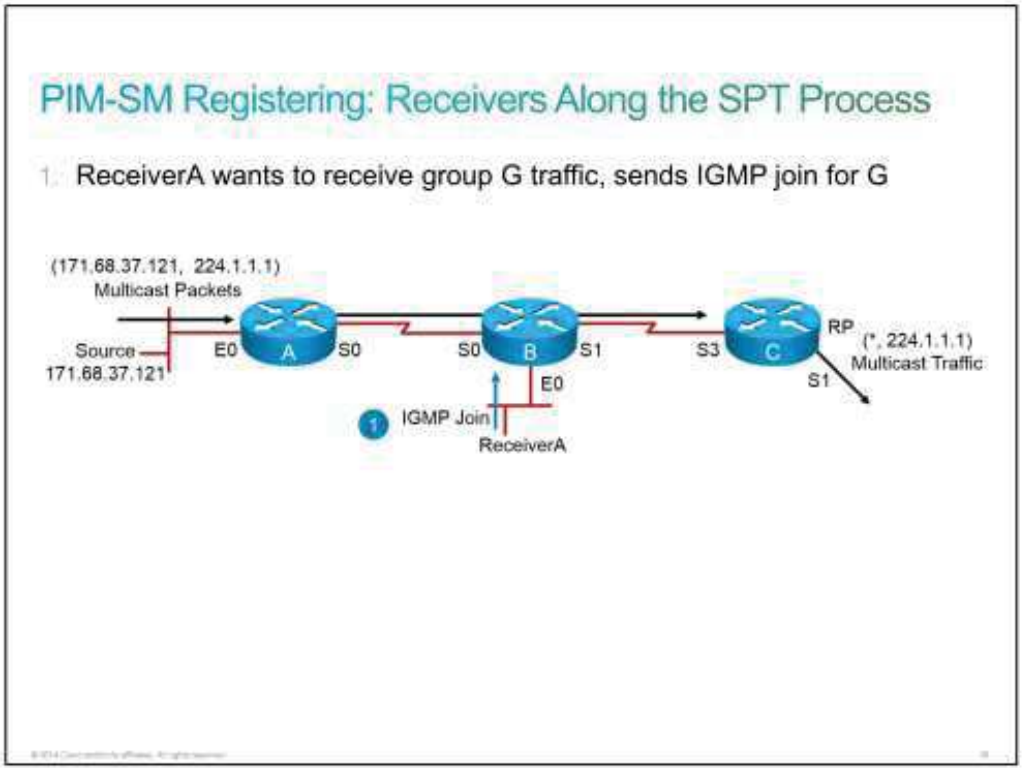

When inspecting the state in the RP with traffic flowing on the SPT, note the following:

- The (*,G) entry only has S1 in its OIL.
- In the (S,G) entry, S3 is the incoming interface, because the RPF interface for source S is via router B.
- S1 is listed in the OIL of the (S,G) entry, because the OIL of the (S,G) entry is always kept in synchronization with the (*,G) OIL.

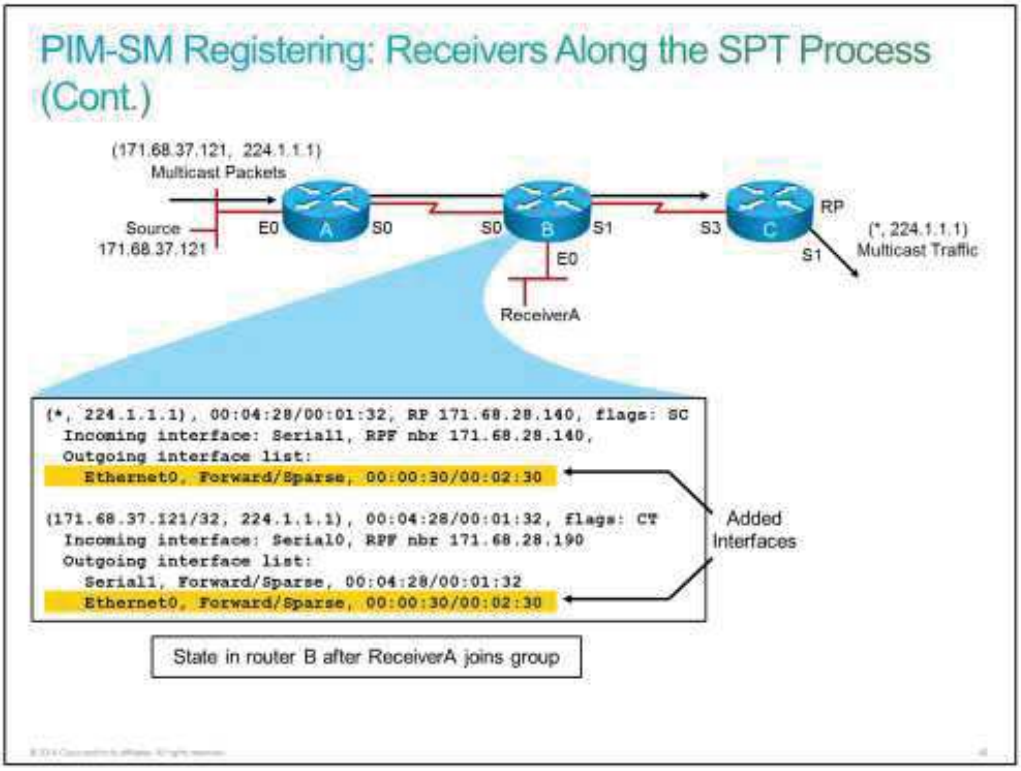
 © 2014 Cisco Systems, Inc.

https://t.me/learningnets

# PIM-SM Registering: Receivers Along the SPT Process




1. A host that is directly connected to router B (ReceiverA) joins the multicast group 224.1.1.1 by sending an IGMP report.



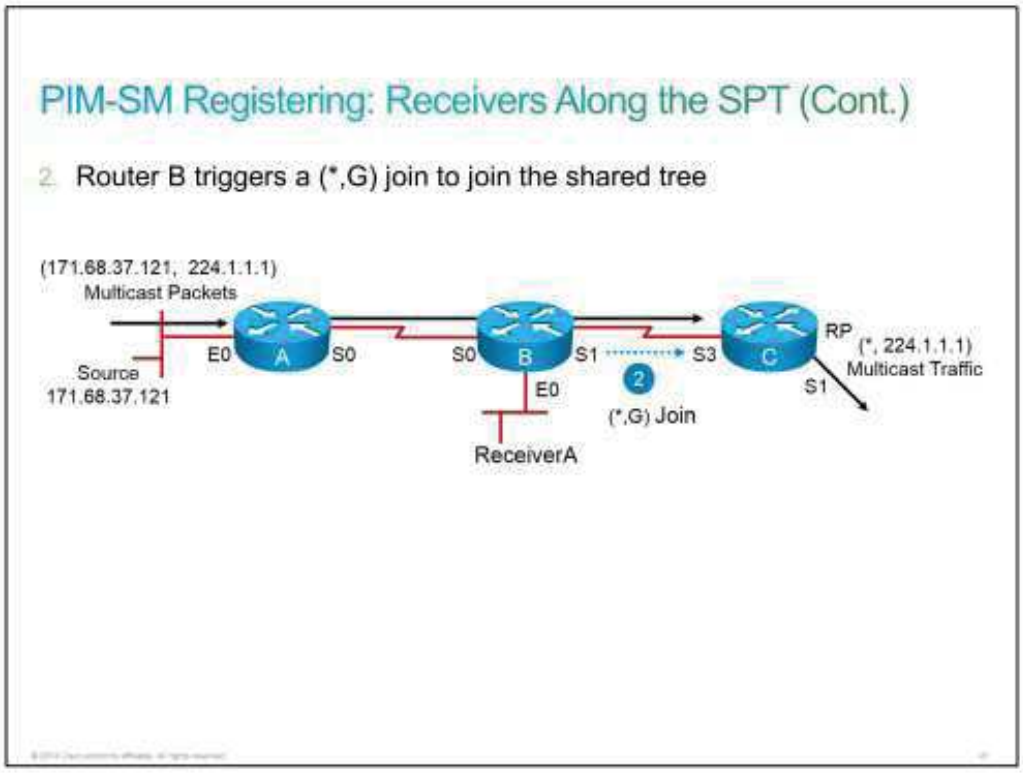



https://t.me/learningnets

In response to the receipt of the IGMP report sent by ReceiverA for group 224.1.1.1, router B updates its state for this group as follows:

- E0 is added to the OIL of the (*,G) entry. This addition is made to permit any (*, 224.1.1.1) traffic flowing down the shared tree to be forwarded to ReceiverA.
- Next, the OILs of all child (S,G) entries are synchronized with the change that was just made to the OIL of the (*,G). This action results in E0 being added to the OIL of the (171.68.37.121/32, 224.1.1.1) entry. This action further permits traffic from this source to be picked off as it flows along the SPT through router B on its way to the RP.

The multicast traffic does not flow to the RP and then back out the same interface to reach routers along the SPT (for example, router B). This belief is a common misperception. The traffic is diverted directly to receiver segments, even if receivers are not directly connected to the routers along the SPT. In the latter case, the (S,G) entries are used first for forwarding and then the (*,G) entries are used.

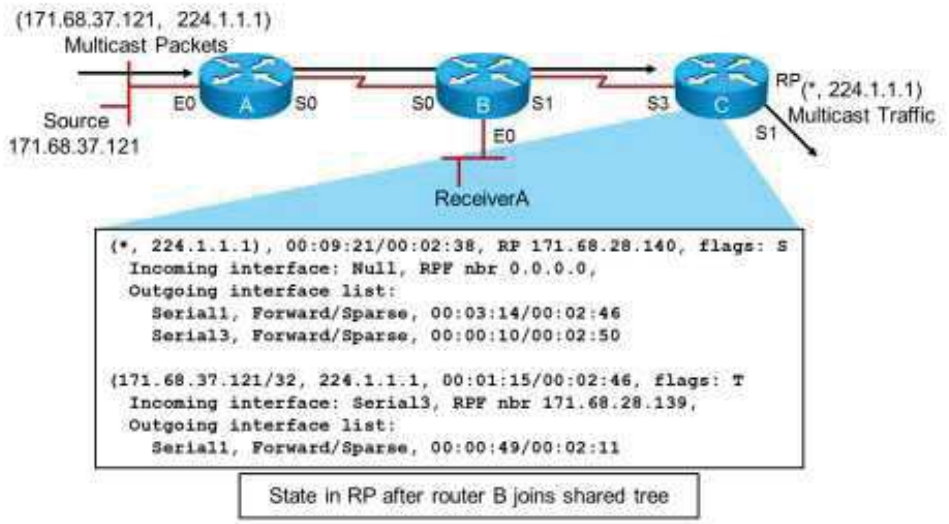



2. Because the OIL of the (*,G) entry in router B transitioned from null to non-null (E0 was added to the [*,G] OIL), a (*,G) join message is triggered. This message is sent up the shared tree toward the respective RP so that router B is placed on a branch of the shared tree. This action is taken so that traffic from other sources to group G will flow down this new branch of the shared tree to reach the receiver.



https://t.me/learningnets

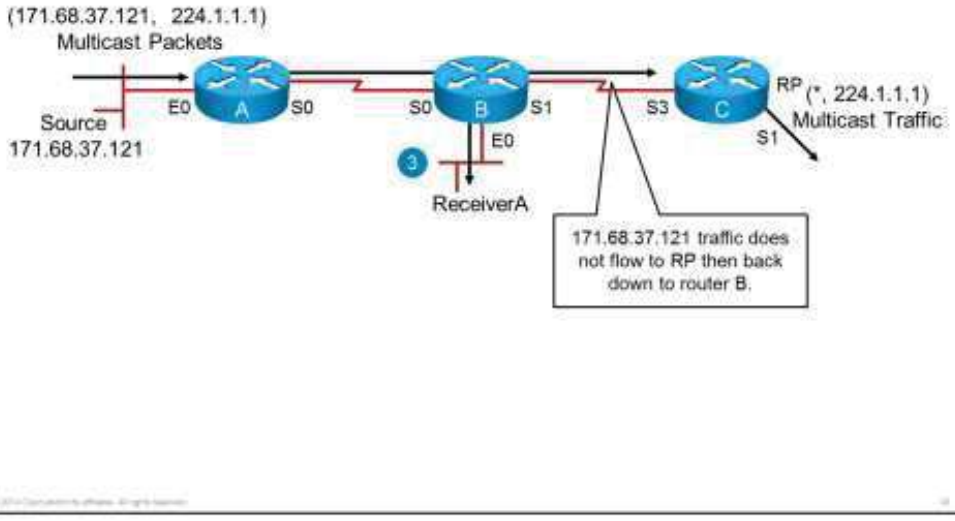
## PIM-SM Registering: Receivers Along the SPT Process (Cont.)




When the RP receives the (*,G) join sent by router B, it adds S3 to the (*,G) OIL. The RP synchronizes the OILs of all (S,G) entries by adding S3 to each (S,G) OIL. However, in this case, S3 is the incoming interface for the (171.68.37.121/32, 224.1.1.1) entry and is therefore not added to the OIL of the (S,G) entry. If it were added, S3 would appear in both the incoming and OILs, which may cause a routing loop. The (S,G) entries on the SPT contain the OIL from the parent (*,G) without the IIF (S3) toward the source S.

## PIM-SM Registering: Receivers Along the SPT Process (Cont.)

3. Group G traffic begins to flow to ReceiverA






https://t.me/learningnets

3. Traffic from source 171.68.37.121 is now being picked off by router B and forwarded out E0 as the traffic flows down the SPT to the RP. It is important to note that this source traffic does not flow to the RP and then return on the same interface on which it arrived. It is diverted directly to the receiver segments. The traffic still flows toward the RP, because there are receivers on the other branches of a shared tree behind the RP.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# PIM-SM SPT Switchover Overview

In PIM-SM, SPT thresholds may be configured to control when traffic switches from the shared tree to the SPT. The decision to switch to the SPT may be done in last-hop routers.

## PIM-SM SPT Switchover Overview

PIM-SM SPT switchover characteristics:

- SPT thresholds may be set for any group:
  - Access lists may be used to specify which groups.
  - Default threshold = 0 kb/s (i.e., immediately join SPT).
  - Threshold of infinity means never join SPT.
- Threshold triggers join of source tree:
  - Pro: Reduces network latency.
  - Con: More (S,G) state must be stored in the routers.

## PIM-SM SPT Switchover Overview (Cont.)

SPT switchover mechanism:

- Once each second:
  - Compute new (*,G) traffic rate
  - If threshold exceeded, set J flag in (*,G)
- For each ($S_i$,G) packet received:
  - If J flag set in (*,G):
    - Join SPT for (S*i*,G)
    - Mark ($S_i$,G) entry with J flag
    - Clear J flag in (*,G)



https://t.me/learningnets

SPT thresholds are specified in kilobits per second (kb/s) and may be used with the access list to specify to which groups the threshold applies. The default SPT threshold is 0 kb/s. This default means that all sources are immediately switched to the SPT. If an SPT threshold of infinity is specified for a group, the sources will not be switched to the SPT and will remain on the shared tree.

When the group SPT threshold is exceeded in a last-hop router, the next received packet for the group will cause an (S,G) join to be sent toward the source of the packet. This action builds an SPT from source S to the last-hop router.

A benefit of switching to the SPT is that the most optimal path is (usually) used to deliver the multicast traffic. Depending on the location of the source in relation to the RP, this switch to the SPT may substantially reduce network latency.

The problems of the SPT switch may be seen in networks with several senders, where an increased amount of state must be kept in the routers. In some cases, an infinity threshold may be used to force certain groups to remain on the shared tree when latency is not an issue.

SPT threshold is a frequently misunderstood mechanism. Many people think that the traffic rates of the sources in the group are monitored and compared against the SPT threshold. This belief is not true, because keeping statistics on each source is virtually the same as creating an (S,G) state. Instead, the total aggregate rate of group traffic flowing down the shared tree is calculated once per second. If this total aggregate rate is exceeded, then the next group packet received will cause that source to be switched to the SPT.

Once every second, the aggregate (*,G) traffic rate is computed and checked against the SPT threshold. If the aggregate rate of all group traffic flowing down the shared tree exceeds the threshold, then the J flag is set in the (*,G) entry.

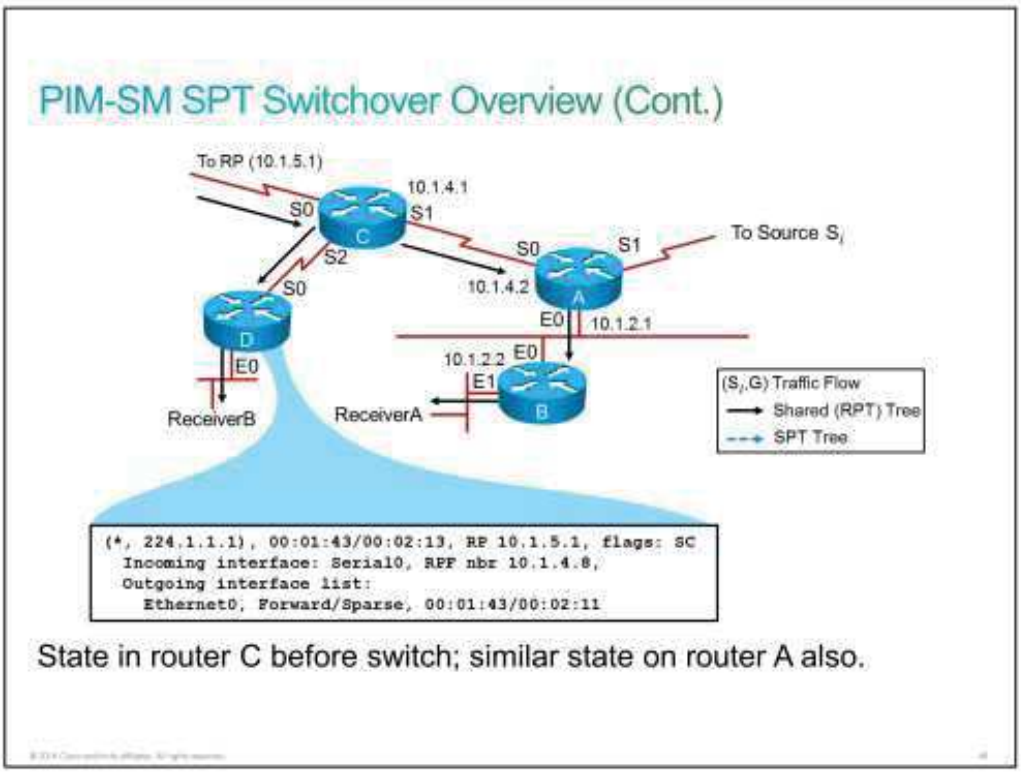
As each multicast packet is received on the shared tree, the J bit is checked in the (*,G) entry.

- If the J flag is set, a new (S,G) entry is created for the source of the packet.
- An (S,G) join is sent toward the source to join the SPT.
- The J flag is set in the (S,G) entry to denote that this entry was created as a result of the SPT threshold switchover.
- The J flag in the (*,G) is reset. It will be set in 1 second if the aggregate rate on the shared tree is still over the SPT threshold.

This mechanism may sometimes result in low-rate sources being switched to the SPT erroneously. However, the RPT-switchback mechanism will correct this situation and, eventually, only the high-rate sources will be received via SPTs while low-rate sources will remain on the shared tree.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

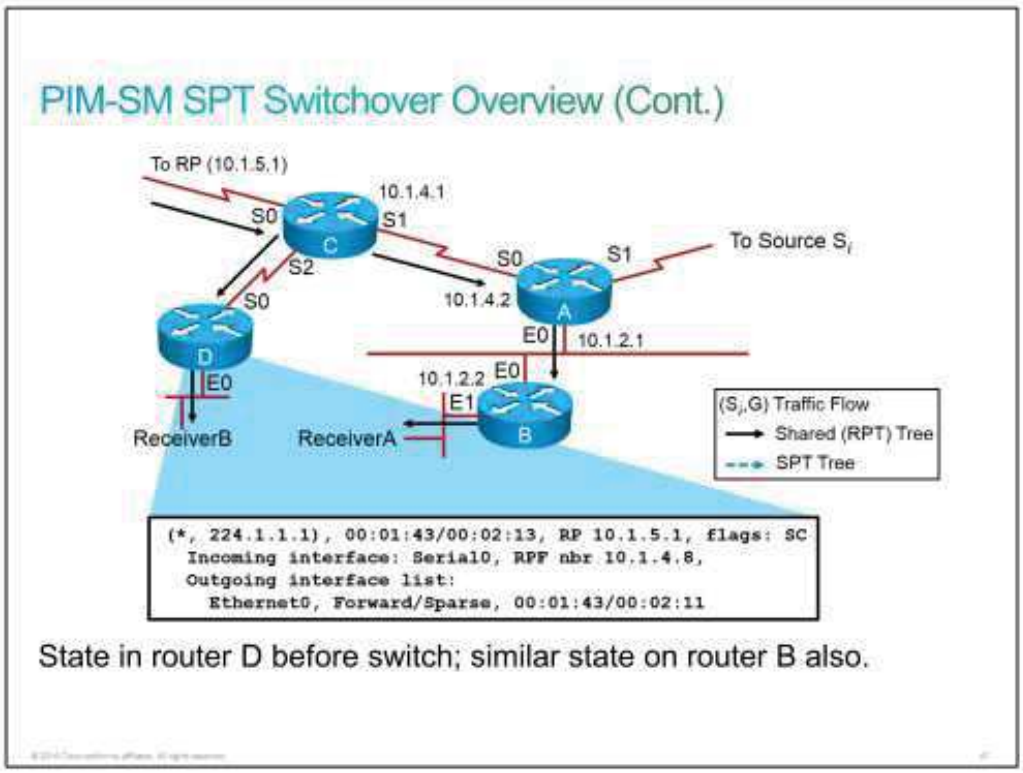



This figure explains the PIM-SM SPT switchover. Note the initial state shown in the figure and observe the following:

- ReceiverA and ReceiverB have joined multicast group 224.1.1.1, which has resulted in traffic flowing down the shared tree, as shown by the solid arrows.
- The state in router C before the switchover, which is as follows:
  - The IIF of the (*,G) entry points toward the RP via S0.
  - The OIL of the (*,G) entry contains S1 and S2 because of (*,G) joins that were sent up the shared tree by routers A and D, respectively.
- The state in router A is similar to the state in router C. The state contains the (*,G) entry only, with one interface in the OIL (the interface via which the [*,G] join was propagated from the last-hop router upstream toward the RP).



https://t.me/learningnets




The state in router D before the switchover, which is as follows:

- The IIF of the (*,G) entry points toward the RP via S0.
- The OIL of the (*,G) entry contains E0 because of the IGMP reports for group 224.1.1.1 that are sent by ReceiverB.

The state in router B before the switchover, which is similar to the state in router D and is as follows:

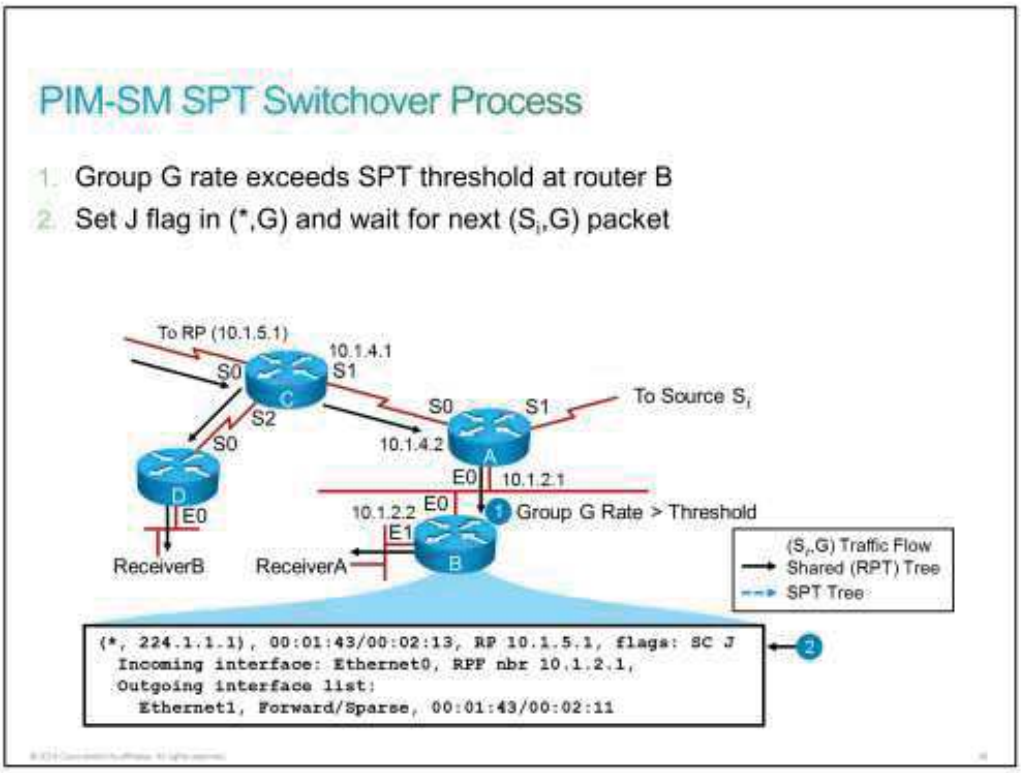
- The IIF of the (*,G) entry points toward the RP via E0.
- The OIL of the (*,G) entry contains E1 because of the IGMP reports for group 224.1.1.1 that are sent by ReceiverA.



https://t.me/learningnets

# PIM-SM SPT Switchover Process

The PIM-SM SPT switchover process is described below:




1. The total amount of all traffic flowing down the shared tree begins to exceed the SPT threshold that is configured at router B (the last-hop router with directly attached receivers of the group G).
2. The router sets the J flag in the (*,G) entry to denote that the rate is above the SPT threshold for this group.

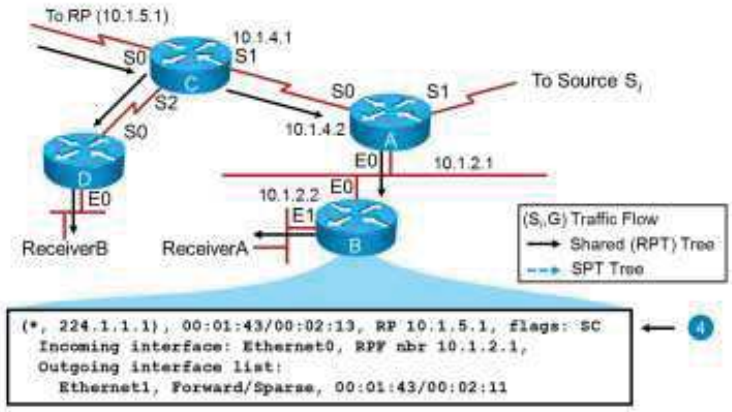
The J flag in the (*,G) entry may be read as the candidate for SPT switchover.



https://t.me/learningnets

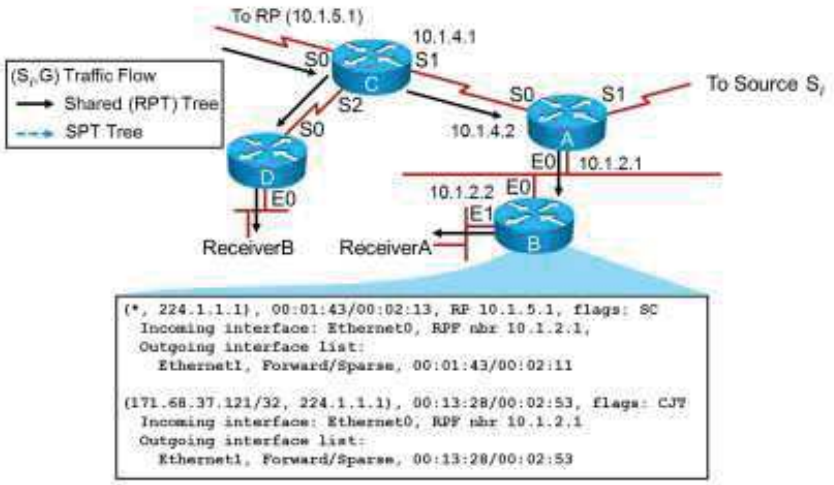
## PIM-SM SPT Switchover Process (Cont.)

3. ($S_i$,G) packet arrives down shared tree
4. Clear J flag in the (*,G) and create ($S_i$,G) state




3. The very next packet to arrive via the shared tree is from source Si for the group G. Because there is a member that is directly connected to router B (denoted by the C flag), and the aggregated group traffic rate is above the SPT threshold (denoted by the J flag), router B initiates a switch to the SPT for ($S_i$,G) traffic.
4. The J flag in the (*,G) entry is first cleared and a new traffic rate measurement interval (1 second) is started. Next, the ($S_i$,G) state is created for source S, sending multicast stream to group G.

## PIM-SM SPT Switchover Process (Cont.)



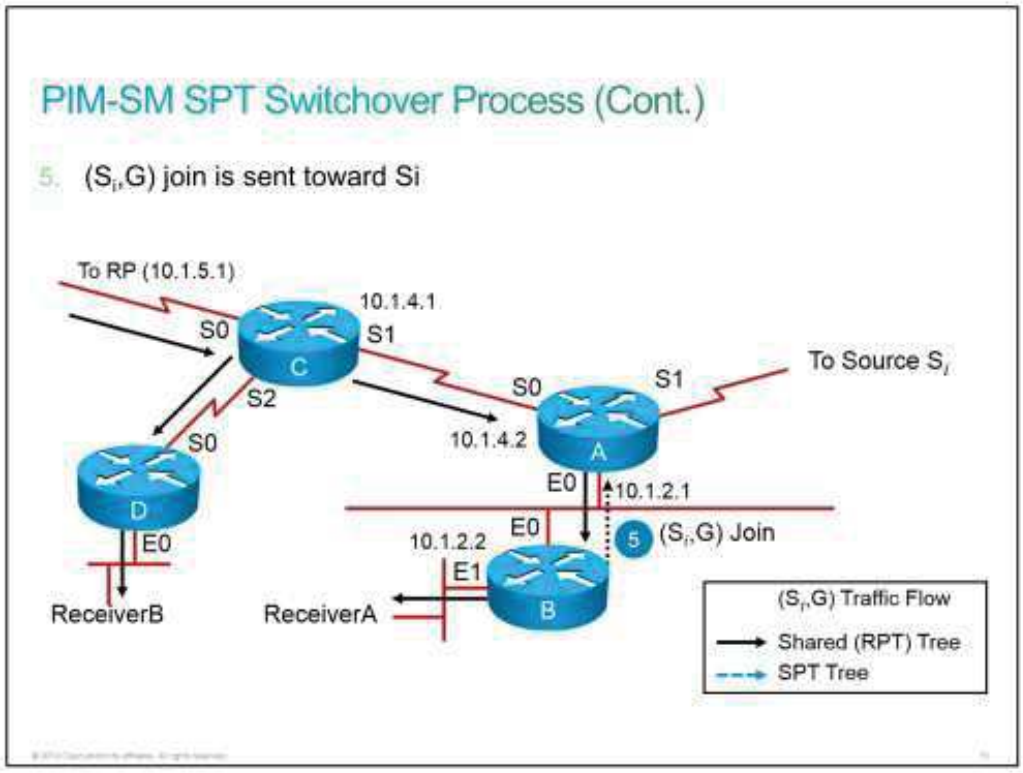

New state in router B.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

The (171.68.37.121/32, 224.1.1.1) entry is created as follows:

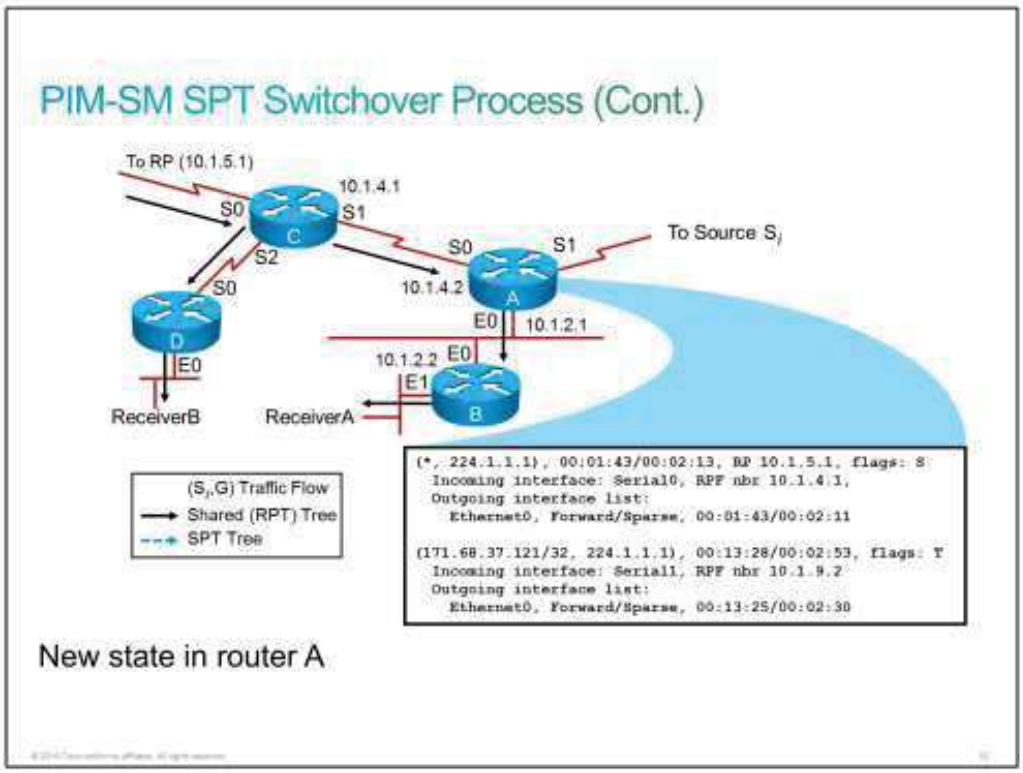
- To denote that this entry was created as a result of the SPT switchover mechanism, the J flag is set on the (S,G) entry. (The setting of the J flag causes router B to monitor the rate of the [S,G] traffic. If the average 1-minute rate of this traffic drops below the SPT threshold, router B will attempt to switch this traffic flow back to the shared tree.)
- RPF information is calculated in the direction of source *Si*. This action results in an IIF of E0, and an RPF neighbor address of 10.1.2.1. (Note that the RPF information for the [S,G] entry is the same as the [*,G] entry. This data indicates that the shared tree and the SPT are following the same path at this point.)
- The OIL for the (S,G) entry is constructed by copying the OIL from the (*,G) entry and then removing the IIF from this list to prevent a possible route loop. This action results in an (S,G) OIL containing only E1.




5. When a state has been created for ($S_i$,G), an ($S_i$,G) join is sent toward source Si to build a branch of the SPT to router B. These ($S_i$,G) joins will continue to be sent periodically (once per minute) as long as the ($S_i$,G) entry is not pruned (that is, it does not have a null OIL).



https://t.me/learningnets




When the ($S_i$,G) join is received by router A, the (171.68.37.121/32, 224.1.1.1) entry in this figure is created as follows:

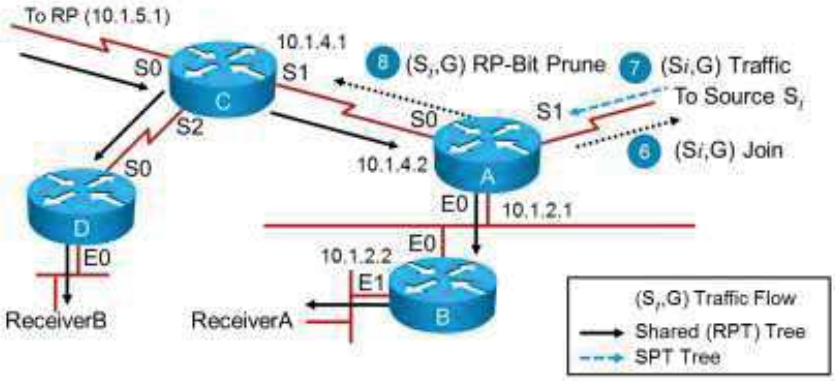
- RPF information is calculated in the direction of source *Si*. This action results in an IIF of S1 and an RPF neighbor address of 10.1.9.2. The RPF information for the (S,G) entry is not the same as the information for the (*,G) entry. This difference indicates that the paths of the shared tree and the SPT diverge at this point.
- The OIL for the (S,G) entry is constructed by copying the OIL from the (*,G) entry and then removing the IIF from this list to prevent a possible route loop. This action results in an (S,G) OIL containing only E0.



https://t.me/learningnets

## PIM-SM SPT Switchover Process (Cont.)

6. Router A forwards ($S_i$,G) join toward Si
7. Packets start arriving down the source tree.
8. SPT and shared tree diverge, triggering ($S_i$,G) RP-bit prunes toward RP

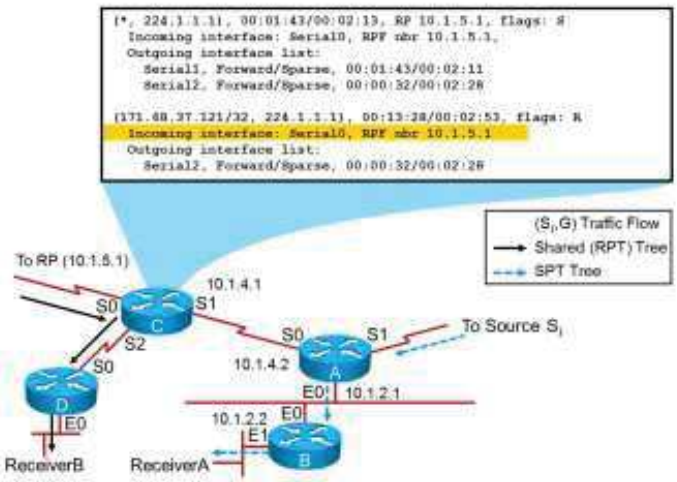



6. When the ($S_i$,G) state is created at router A, an ($S_i$,G) join is sent toward source S. These ($S_i$,G) joins continue to be sent periodically (once per minute) as long as the ($S_i$,G) entry is not pruned (that is, it does not have a null OIL).
7. Packets start arriving down the source tree. When the ($S_i$,G) joins reach the first-hop router that is directly connected to source S*i*, a complete branch of the SPT has been built (shown in the figure by the label 7 dashed arrow line). This branch permits ($S_i$,G) traffic to flow directly via the SPT to router B and ReceiverA.
8. Because the paths of the shared tree and the SPT diverge at router A (there is a difference in RPF information between the [*,G] and [$S_i$,G] entries), router A begins to send ($S_i$,G) RP-bit prune messages up the shared tree to stop the flow of redundant ($S_i$,G) traffic down the shared tree.



https://t.me/learningnets

## PIM-SM SPT Switchover Process (Cont.)




State in router C after receiving the (S*i*,G) RP-bit prune.

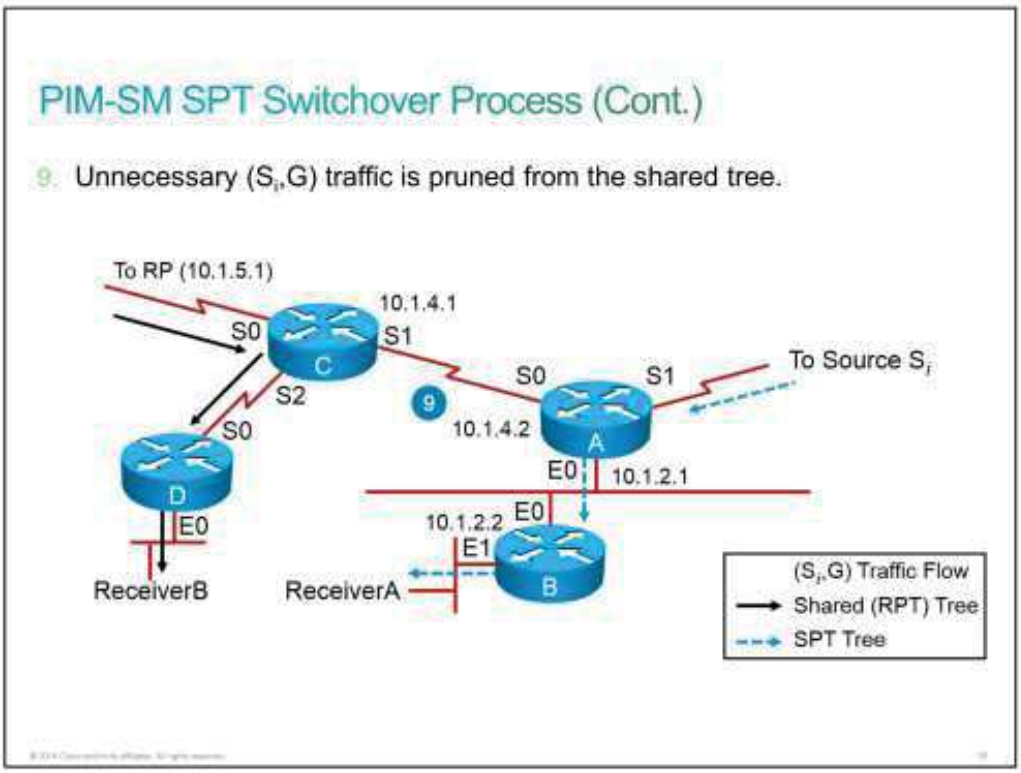
When the ($S_i$,G) RP-bit prune reaches router C, the (171.68.37.121/32, 224.1.1.1) entry in this figure is created as follows:

- Because this ($S_i$,G) entry was created in response to the receipt of an ($S_i$,G) RP-bit prune, the R bit is set to denote that this forwarding state is applicable to traffic flowing down the shared tree and not the source tree.
- Because the R bit is set, the RPF information is calculated in the direction of the RP instead of source S*i*. (This entry is applicable to [$S_i$,G] traffic flowing down the shared tree, and therefore the RPF information must point up the shared tree.) This action results in an IIF of S0, and an RPF neighbor address of 10.1.5.1.
- The OIL for the ($S_i$,G) entry is constructed by copying the OIL from the (*,G) entry without the interface that the ($S_i$,G) RP-bit prune was received. Next, the IIF is removed from the OIL to prevent a possible route loop. These steps result in an ($S_i$,G) OIL containing only S2.

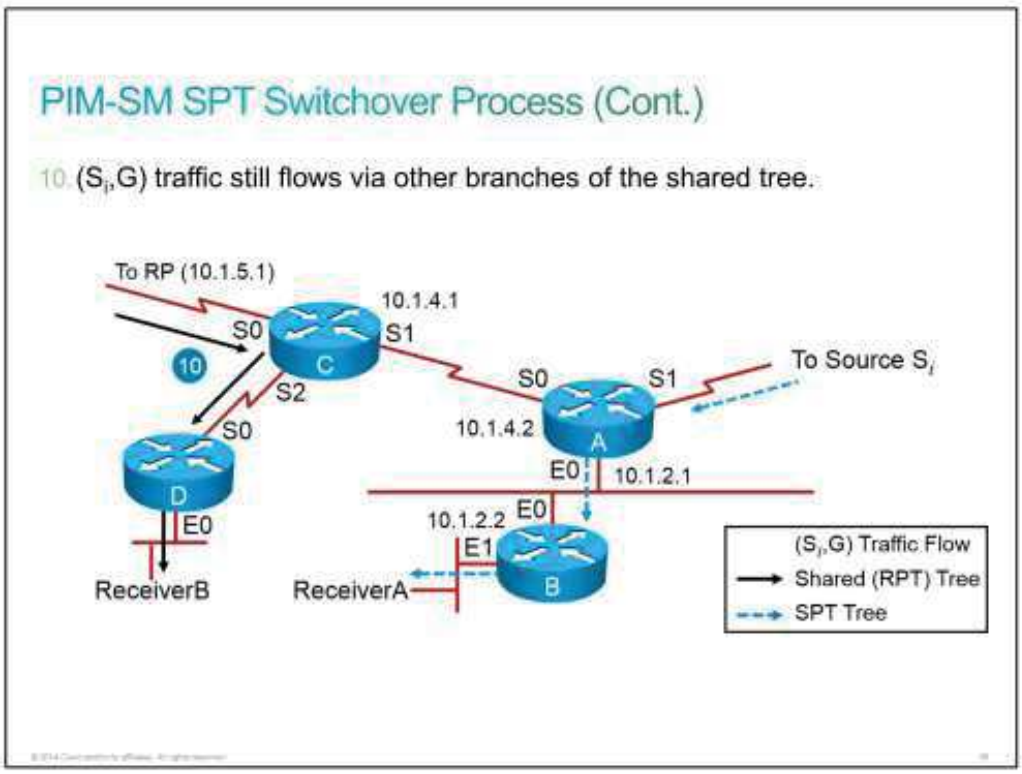
At this point, ($S_i$,G) traffic flowing down the shared tree is forwarded using the ($S_i$,G) entry. The ($S_i$,G) traffic arriving at router C will pass the RPF check, because the RPF information in the ($S_i$,G) entry is pointing up the shared tree (because of the R bit). The traffic will then be forwarded to all interfaces in the ($S_i$,G) OIL. In this case, only S2 remains in the ($S_i$,G) OIL, and therefore ($S_i$,G) traffic will be sent to router D but not router A. This activity successfully prunes the redundant ($S_i$,G) traffic from the branch of the shared tree between router C and router A.



https://t.me/learningnets




9. At this point, the redundant (S,G) traffic is pruned from the shared tree branch from router C to router A. (S,G) traffic is reaching ReceiverA via the SPT through routers A and B directly from the source segment.



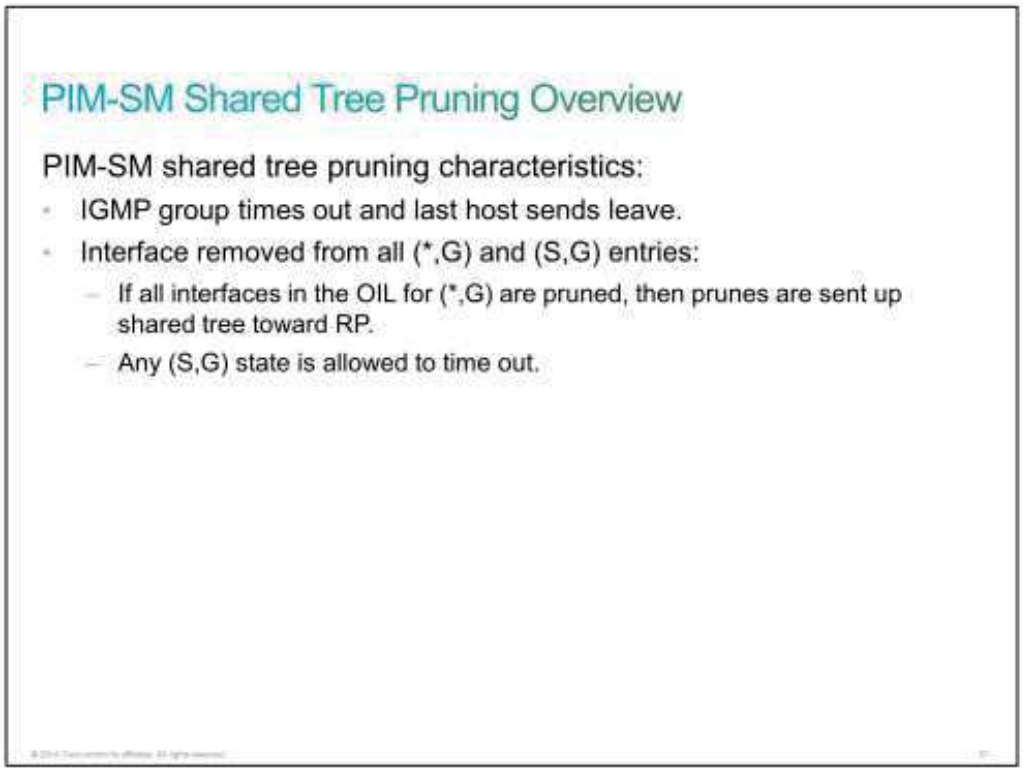

10. (S,G) traffic is still reaching ReceiverB via a branch of the shared tree through routers C and D (although using the [S,G] entry). This activity occurs because the (S,G) state in router C still has S2 in its OIL.



https://t.me/learningnets

# PIM-SM Shared Tree Pruning Overview

When a locally connected host sends an IGMP leave (or an IGMP state times out in the router, if IGMPv1 is used) for group G, the interface is removed from the (*,G) and from all (S,G) entries in the multicast routing table.



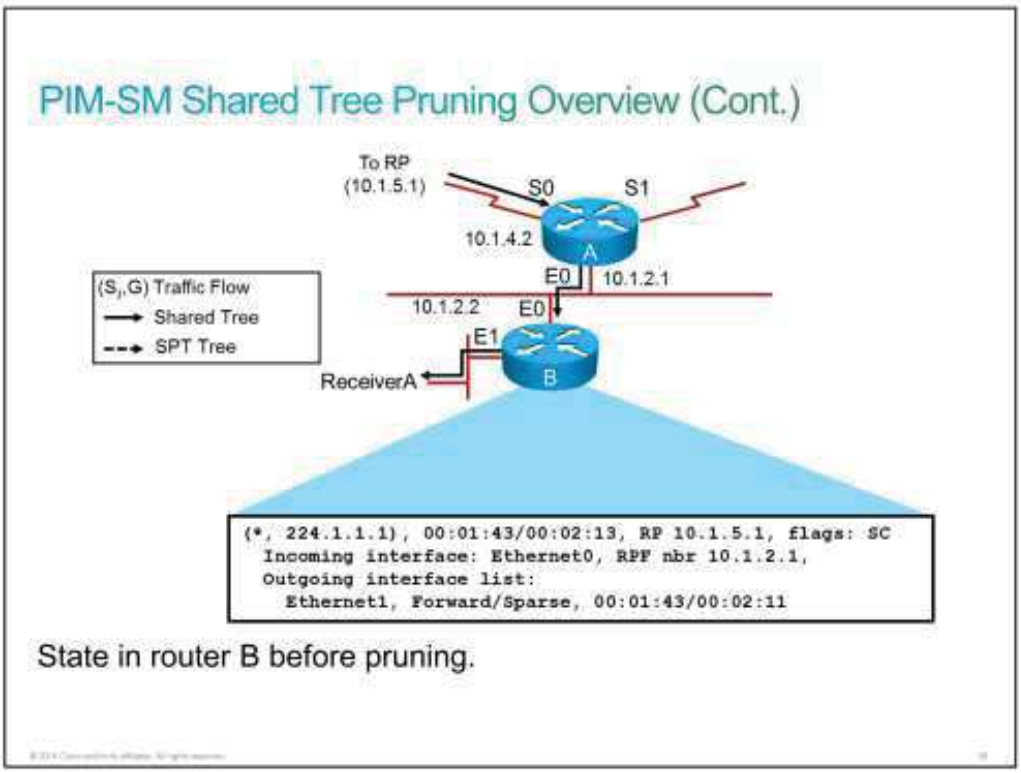

If the (*,G) OIL is now null, then the last-hop DR sends a (*,G) prune up the shared tree toward the RP. Any remaining (S,G) entries are allowed to time out and be deleted from the multicast routing table.

When the routers up the shared tree receive the (*,G) prune, they remove the interface on which the prune was received from their (*,G) OIL.

If, as a result of removing the interface, the (*,G) OIL becomes null, then the routers forward a (*,G) prune up the shared tree toward the RP. Any remaining (S,G) entries are allowed to time out and be deleted from the multicast routing table.



https://t.me/learningnets



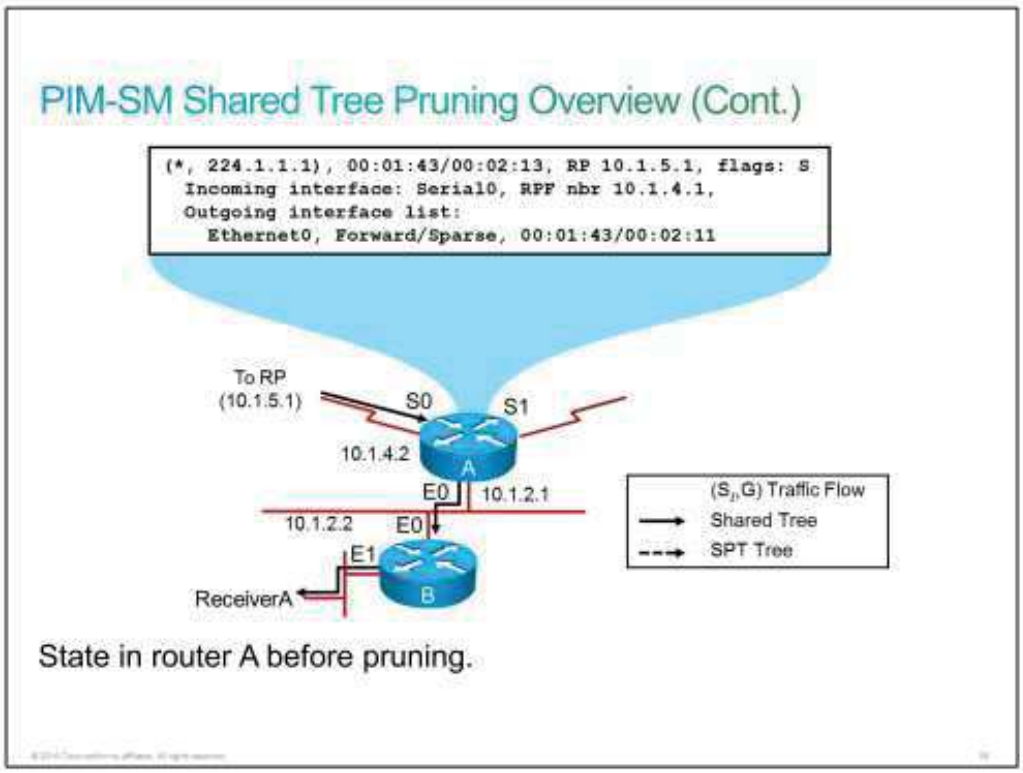

In this figure, the traffic is flowing down the shared tree (denoted by the existence of only the [*,G] entry in the multicast forwarding table in router B).

| | |
|---|---|
| **Note** | The (*,G) entry in the last-hop router does not necessarily mean that the traffic is actually flowing down the shared tree. The entry only indicates that there are active group members and that the branch of a shared tree for the active group was established. For an exact check of the traffic flow, the group counters have to be inspected. |

The incoming interface in the (*,G) entry in router B is E0. The OIL contains E1, and the C flag is set in the (*,G), which denotes that there is a locally connected host for this group (ReceiverA).



https://t.me/learningnets




When inspecting the state in router A before pruning (in the case of a shared tree), note the following:

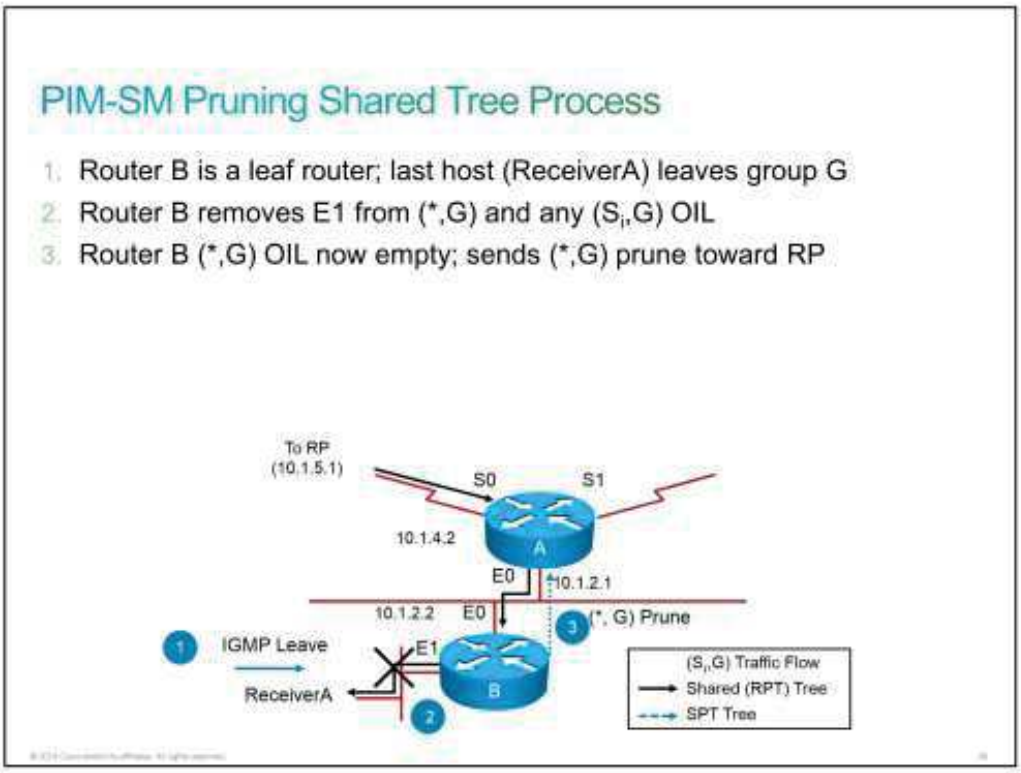
- Traffic is flowing down the shared tree (denoted by the existence of only the [*,G] entry).
- The incoming interface is S0.
- The OIL contains E0.

 © 2014 Cisco Systems, Inc.

https://t.me/learningnets

# PIM-SM Pruning Shared Tree Process

The PIM-SM pruning shared tree process is described below:



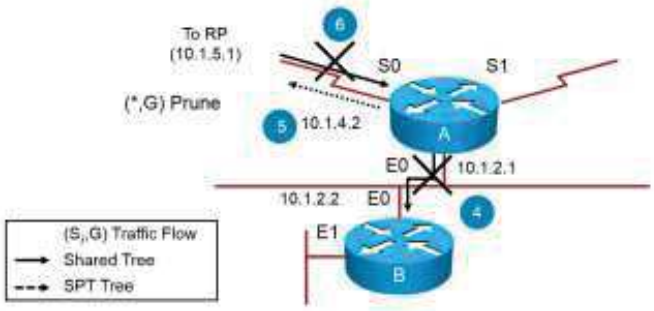

1. When the last-hop router B receives an IGMP group-leave message from ReceiverA for group G, it performs the normal IGMP leave processing. It also finds that ReceiverA was the last host to leave. The IGMP state for group G on interface E1 is deleted.
2. This action causes interface E1 to be removed from the OIL of the (*,G) entry and from any (S,,G) entries (in this case, there are none) in the multicast routing table.
3. Because E1 was the only interface in the (*,G) entry, its OIL becomes null, which results in a (*,G) prune sent up the shared tree via E0 toward the RP.



https://t.me/learningnets

## PIM-SM Pruning Shared Tree Process (Cont.)

4. Router A receives prune, removes E0 from (*,G) OIL (after the 3-second multiaccess network prune delay)
5. Router A (*,G) OIL now empty; sends (*,G) prune toward RP
6. Pruning continues back toward RP



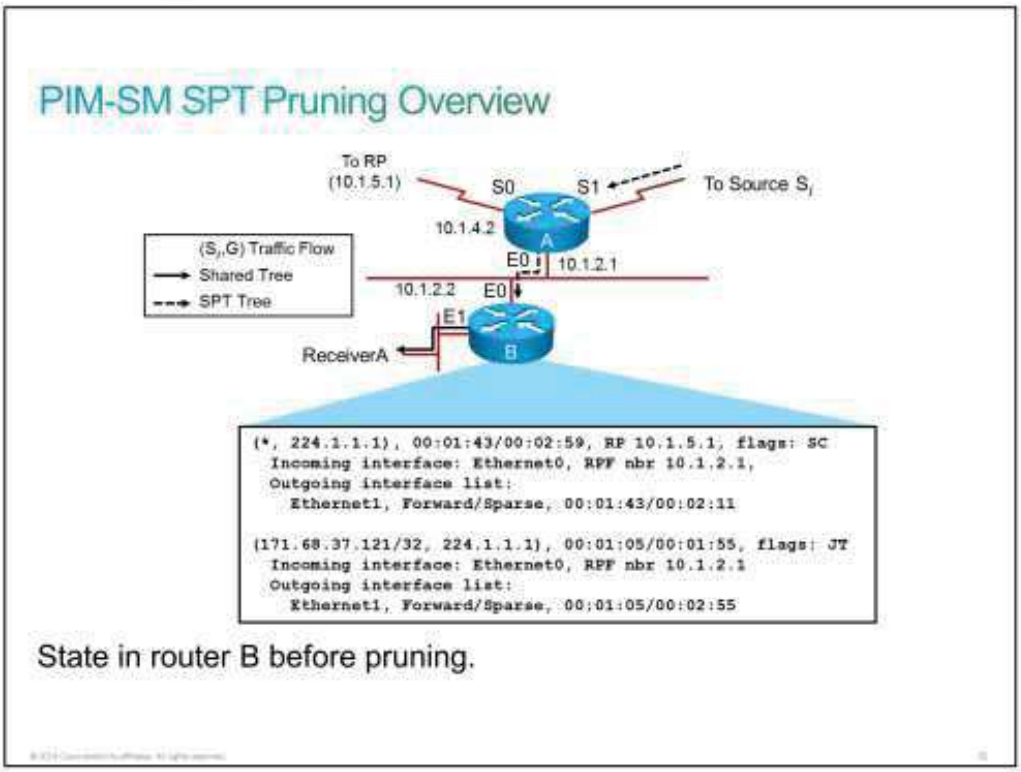

4. Router A will receive a (*,G) prune message and remove interface E0 from the OIL of the (*,G) entry. Router A delayed pruning E0 from the (*,G) entry for 3 seconds because this is a multiaccess network. As such, it needed to wait for a possible overriding join from another PIM neighbor. Because no overriding join was received, the interface was pruned.
5. Because the (*,G) OIL is now null, a (*,G) prune message is forwarded on up the shared tree (RPT) via S0 toward the RP.
6. This pruning continues back toward the RP, or until a router is reached whose (*,G) OIL does not transition to null because of the prune.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# PIM-SM SPT Pruning Overview

The PIM-SM SPT pruning process is described below:




State in router B before pruning.

The following SPT example pertains to pruning on an SPT. In the initial state in router B (the last-hop router) before pruning, note the following:

- Both (*,G) and (S,G) entries exist.
- The J flag is set in the (S,G) entry. This setting indicates that the (S,G) state was created in response to the SPT threshold being exceeded and the switchover being performed.
- The T flag is set in the (S,G) entry. This setting indicates that (S,G) traffic is being successfully received via the SPT.
- The incoming interface is the same for the (*,G) and the (S,G) entries. This similarity indicates that the shared tree and the SPT overlap at this branch of the distribution tree.

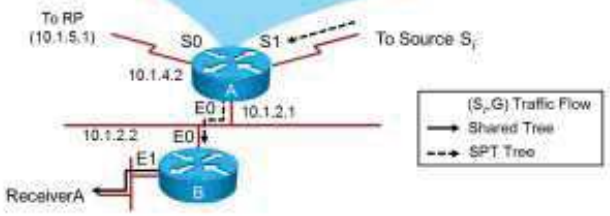


https://t.me/learningnets

## PIM-SM SPT Pruning Overview (Cont.)

```
(*, 224.1.1.1), 00:01:43/00:02:59, RP 10.1.5.1, flags: S
  Incoming interface: Serial0, RPF nbr 10.1.4.1,
  Outgoing interface list:
    Ethernet0, Forward/Sparse, 00:01:43/00:02:11

(171.68.37.121/32, 224.1.1.1), 00:01:05/00:01:55, flags: T
  Incoming interface: Serial1, RPF nbr 10.1.9.2
  Outgoing interface list:
    Ethernet0, Forward/Sparse, 00:01:05/00:02:55
```




State in router A before pruning.

When inspecting the state in router A before pruning (in the case of an SPT), note the following:

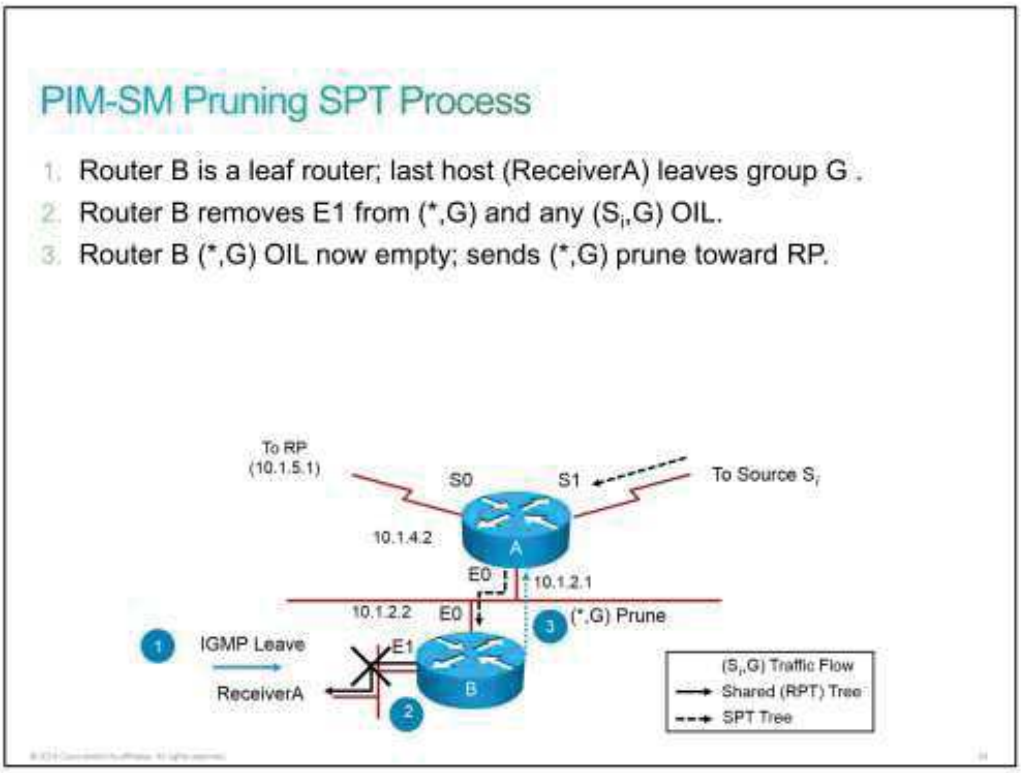
- Both the (*,G) and (S,G) entries exist.
- The T flag is set in the (S,G) entry, which indicates that (S,G) traffic is being successfully received via the SPT.
- The incoming interface is different for the (*,G) and the (S,G) entries. This difference indicates that the shared tree and the SPT diverge at this point (on router A).



https://t.me/learningnets

# PIM-SM Pruning SPT Process

The PIM-SM pruning SPT process is described below:



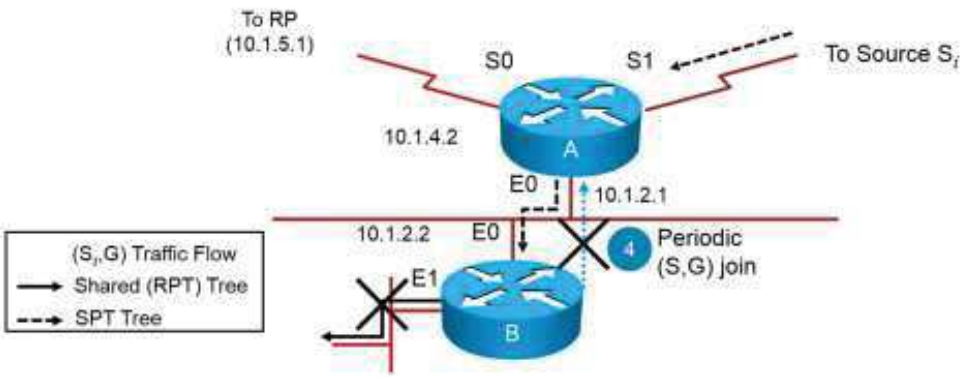

1. When the last-hop router B receives an IGMP group-leave message from ReceiverA for group G, it performs the normal IGMP leave processing and finds that ReceiverA was the last host to leave. The IGMP state for group G on interface E1 is deleted.
2. This action causes interface E1 to be removed from the OIL of the (*,G) entry and from any ($S_i$,G) entries in the multicast routing table. Because E1 was the only interface in the (*,G) and the ($S_i$,G) entries, their OILs become null.
3. Because the (*,G) OIL is now null, a (*,G) prune message is sent up the shared tree via E0 toward the RP.



https://t.me/learningnets

## PIM-SM Pruning SPT Process (Cont.)

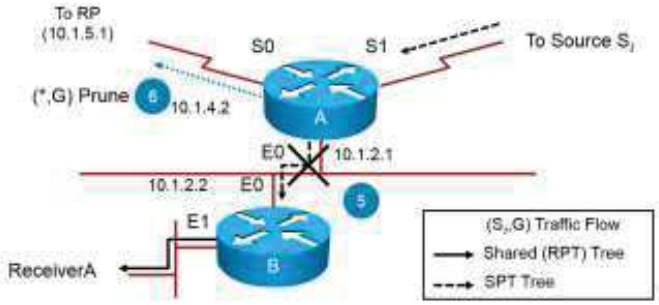
4. Router B stops sending periodic (S,G) joins.




4. Because the ($S_i$,G) OIL is now null, router B stops sending periodic ($S_i$,G) join messages up the SPT. This action results in the removal of the interfaces from the ($S_i$,G) OIL in the upstream router A after a certain amount of time (typically three times the periodic [$S_i$,G] join interval).

## PIM-SM Pruning SPT Process (Cont.)

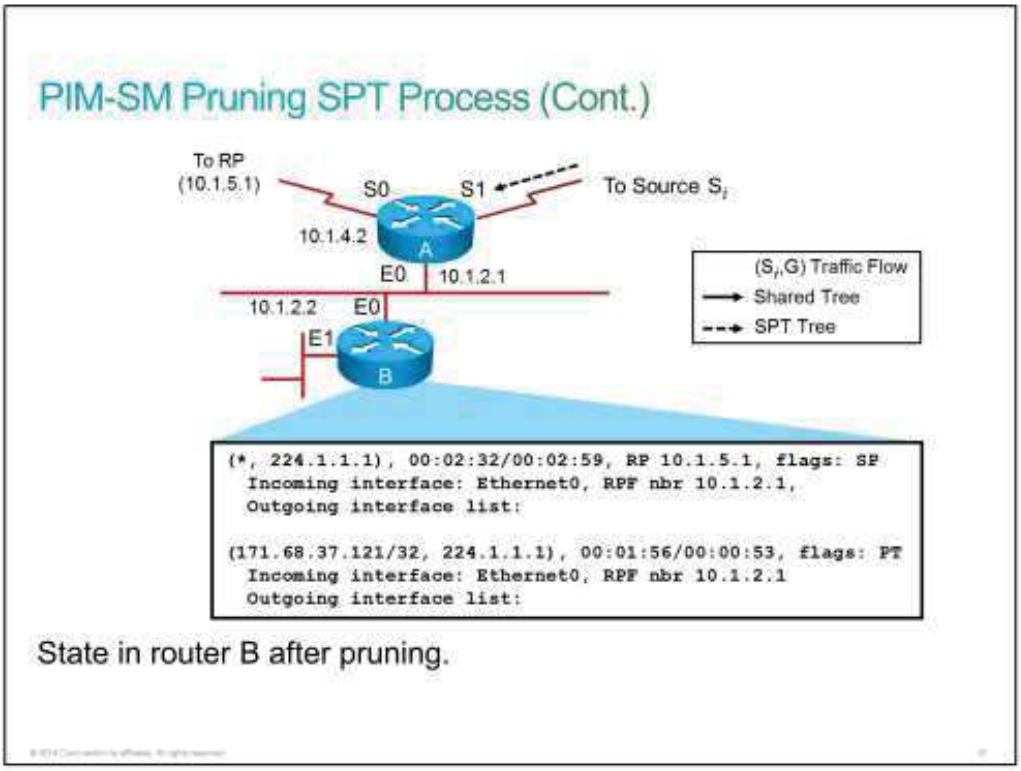
5. Router A receives prune; removes E0 from (*,G) OIL (after the 3-second multiaccess network prune delay).
6. Router A (*,G) OIL now empty; sends (*,G) prune toward RP.




 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

5. Router A will receive a (*,G) prune message and remove interface E0 from the OIL of the (*,G) entry. Router A delayed pruning E0 from the (*,G) entry for 3 seconds because this is a multiaccess network. As such, it needed to wait for a possible overriding join from another PIM neighbor. Because no overriding join was received, the interface was pruned.
6. Because the (*,G) OIL is now null, a (*,G) prune message is sent up the shared tree via S0 toward the RP.




Because the IGMP leave message for group G was received on the E1 interface to the last-hop router B, the interface was removed from both the (*,G) and (S,G) entries. This action caused the OIL to transition to null in both entries, and it set the P flag.

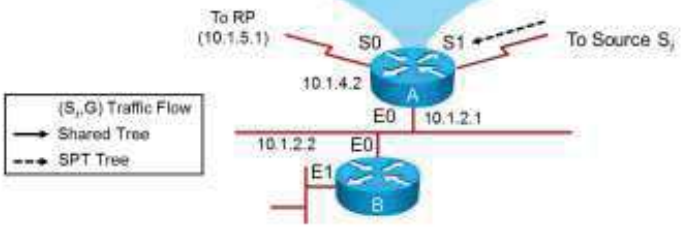


https://t.me/learningnets

## PIM-SM Pruning SPT Process (Cont.)

```
(*, 224.1.1.1), 00:02:32/00:02:59, RP 10.1.5.1, flags: SP
  Incoming interface: Serial0, RPF nbr 10.1.4.1,
  Outgoing interface list:

(171.68.37.121/32, 224.1.1.1), 00:01:56/00:00:53, flags: PT
  Incoming interface: Serial1, RPF nbr 10.1.9.2
  Outgoing interface list:
```



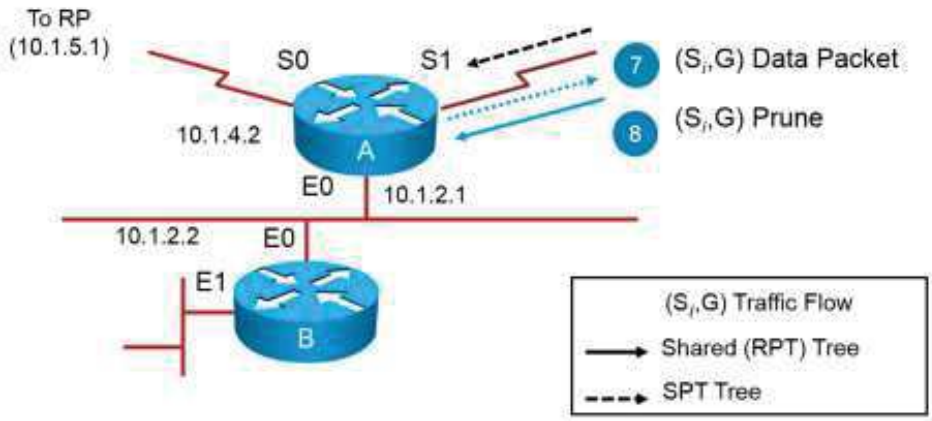

State in router A after pruning.

In response to the (*,G) prune message received by router A, the router removed the E0 interface from the OIL for the (*,G) entry.

Additionally, the interface was removed from the (S,G) entry (the entry contains the copy of the OIL from the parent [*,G] entry). And because periodic (S,G) joins are no longer being sent by the downstream routers, the group expiration timer counts down, and approaches 0.

## PIM-SM Pruning SPT Process (Cont.)

7. Another ($S_i$,G) data packet arrives via Serial1.
8. Router A responds by sending an (Si,G) prune toward source.

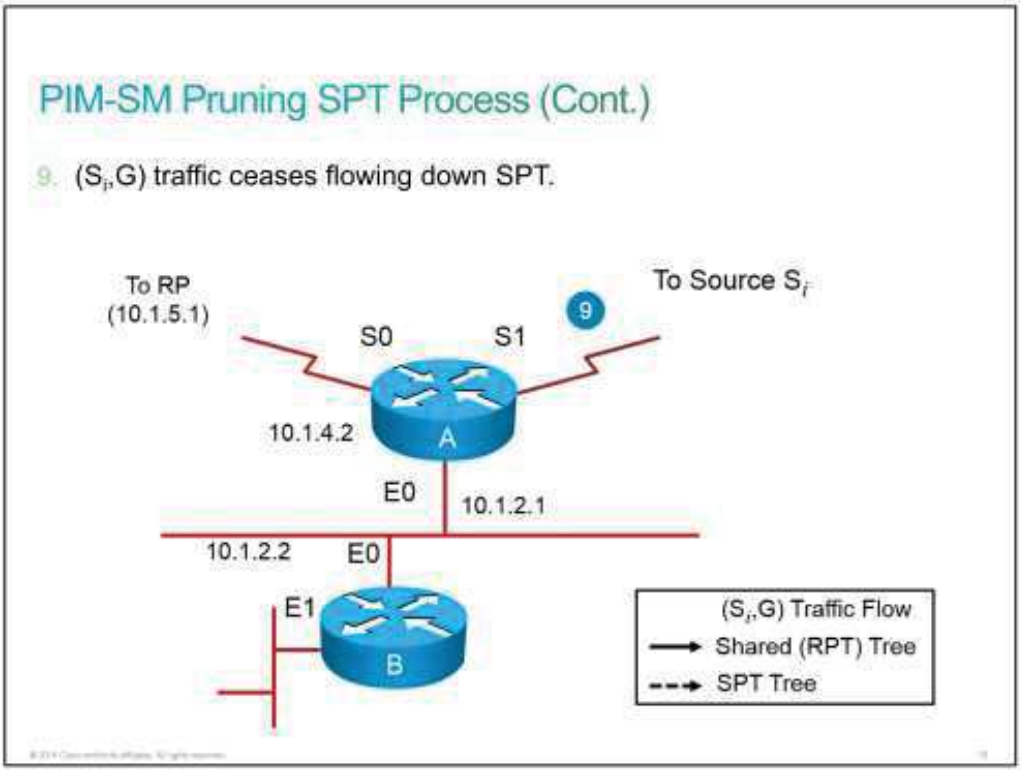



7. Because router A is no longer receiving ($S_i$,G) join messages from router B, the ($S_i$,G) state eventually times out.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

8. This action causes an ($S_i$,G) prune message to be sent up the SPT toward the source Si. The null OIL in the ($S_i$,G) entry does not result in prunes being sent up the SPT. Only periodic ($S_i$,G) joins cease. A prune is sent only if triggered by the traffic that is still flowing down the SPT or by the ($S_i$,G) entry expiration.



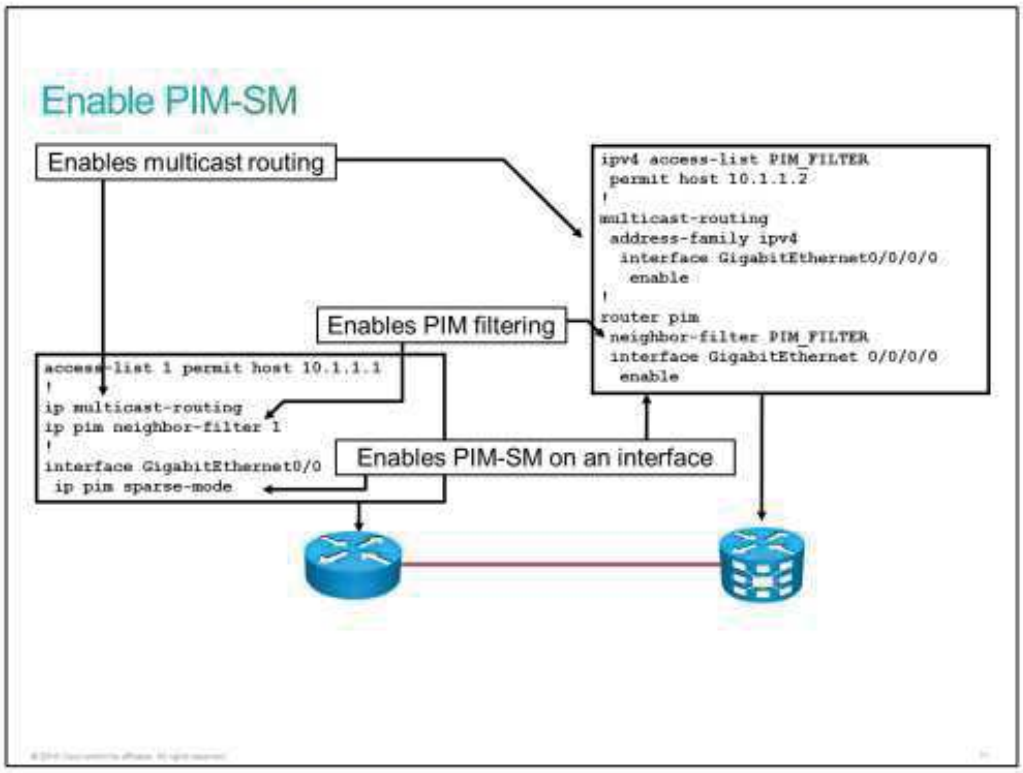

9. After the ($S_i$,G) prune is sent by router A, the traffic stops flowing down the SPT. This result occurs because the upstream router removes the interface (on which the prune was received) from the ($S_i$,G) OIL.



https://t.me/learningnets

# Enable PIM-SM

Basic PIM-SM multicast is very easy to configure.



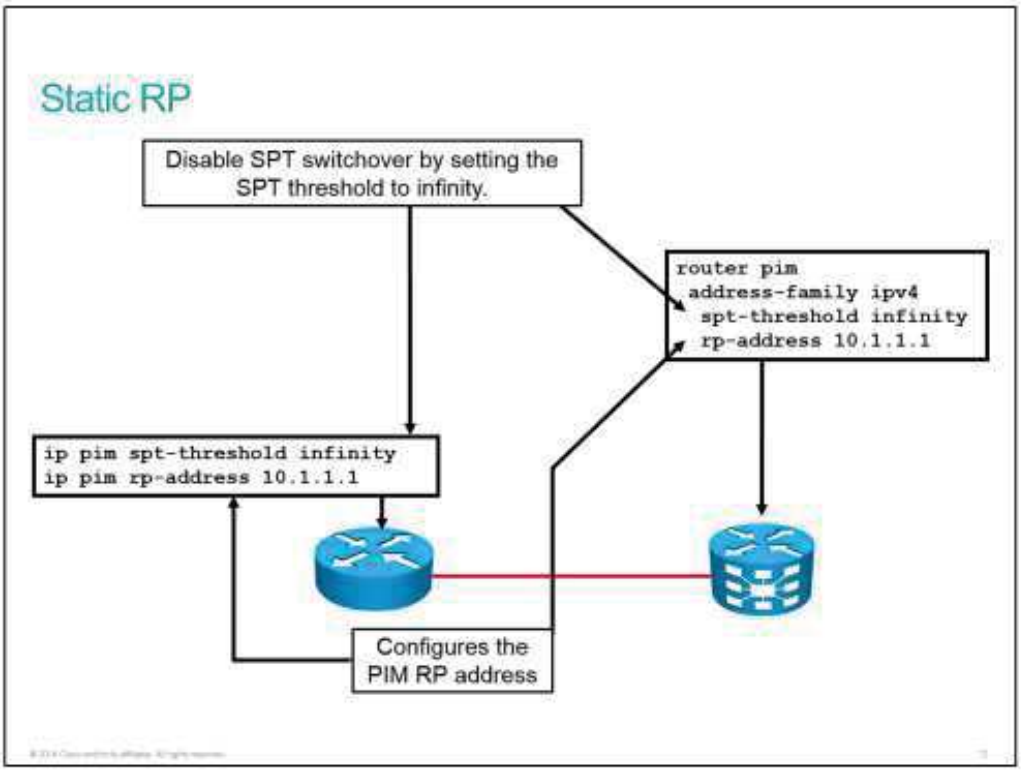

To enable multicast routing, use the Cisco IOS and IOS XE **ip multicast-routing** command or the Cisco IOS XR **multicast-routing** command. Use the Cisco IOS and IOS XE **ip pim sparse-mode** interface command or the Cisco IOS XR **enable** router PIM interface command to enable PIM-SM on the router interface. Use the Cisco IOS and IOS XE **ip pim neighbor-filter** command or the Cisco IOS XR **neighbor-filter** command, followed by an access list, to specify the neighbors that a PIM neighbor relationship can be established with.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# Static RP

Use the Cisco IOS and IOS XE **ip pim rp-address global** command or the Cisco IOS XR **rp-address** router PIM command to configure the IP address of the RP.




An additional command is needed when you do not want to use the Cisco IOS, IOS XE, or IOS XR Software default setting of 0 for the SPT threshold. The default causes the router to switch to the SPT after receipt of the first multicast packet. If this switching is not desired for some or all groups and thresholds are dependent on the group rate, use Cisco IOS and IOS XE **ip pim spt-threshold** global command or the Cisco IOS XR **spt-threshold** router PIM command.

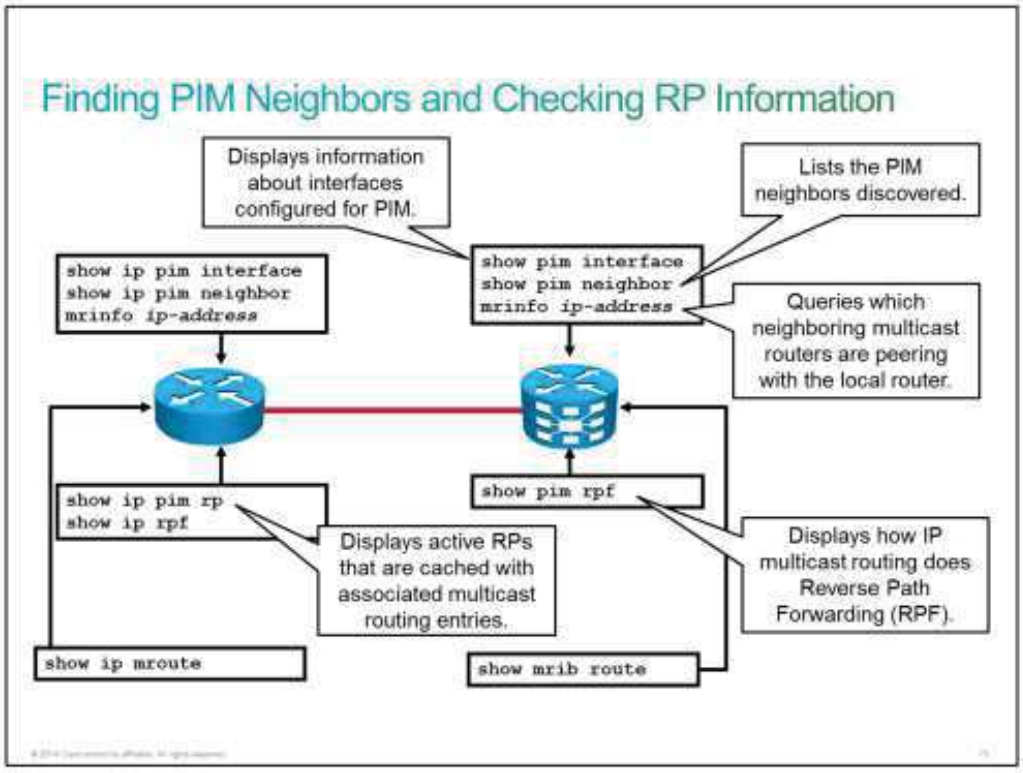
The arguments for the Cisco IOS and IOS XE **ip pim spt-threshold** command or the Cisco IOS XR **spt-threshold** command are the overall group rate (in kbps) and an optional standard access list that defines the affected groups. If the Cisco IOS, IOS XE, or IOS XR **infinity** keyword is used, no switchover is allowed for those groups, and only shared trees will be used.



https://t.me/learningnets

# Finding PIM Neighbors and Checking RP Information

When PIM-SM is configured, the first step is to check proper operation of the network.




Verify that PIM is enabled on the interfaces and determine that the PIM neighbors are correct. The following commands may be used to establish the presence of the PIM neighbors:

- Cisco IOS and IOS XE **show ip pim interface** or Cisco IOS XR **show pim interface**: This command displays the interfaces that are configured for PIM.
- Cisco IOS and IOS XE **show ip pim neighbor** or Cisco IOS XR **show pim neighbor**: This command displays the discovered PIM neighbors.
- Cisco IOS, IOS XE, and IOS XR **mrinfo**: This command displays information on multicast routers that are neighboring with the local router (which has no address) or with the addressed router.

The RP for a certain multicast group operating in sparse mode has to be reachable and known to the router. The troubleshooting of RP information requires (in addition to standard tools like using a unicast ping to check the RP reachability) at least the following three commands:

- Cisco IOS, IOS XE, and IOS XR **show ip pim rp**: The command displays, without arguments, RP information on active groups. If the group address or name is provided, only the RP information for the selected group is shown (assuming that it is an active group).

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

- Cisco IOS and IOS XE **show ip pim rp mapping** or Cisco IOS XR **show pim group-map**: The command displays the contents of the important group-to-RP mapping cache that contains information regarding which RP is active for which group range. The group-to-RP mapping cache is populated via the Auto-RP or BSR mechanism or via static RP assignments. It is very important to check this information to verify that the router possesses the RP mapping information consistent with proper network operation.
- Cisco IOS and IOS XE **show ip rpf** or Cisco IOS XR **show pim rpf**: The command displays RPF information for the RP or for the source.

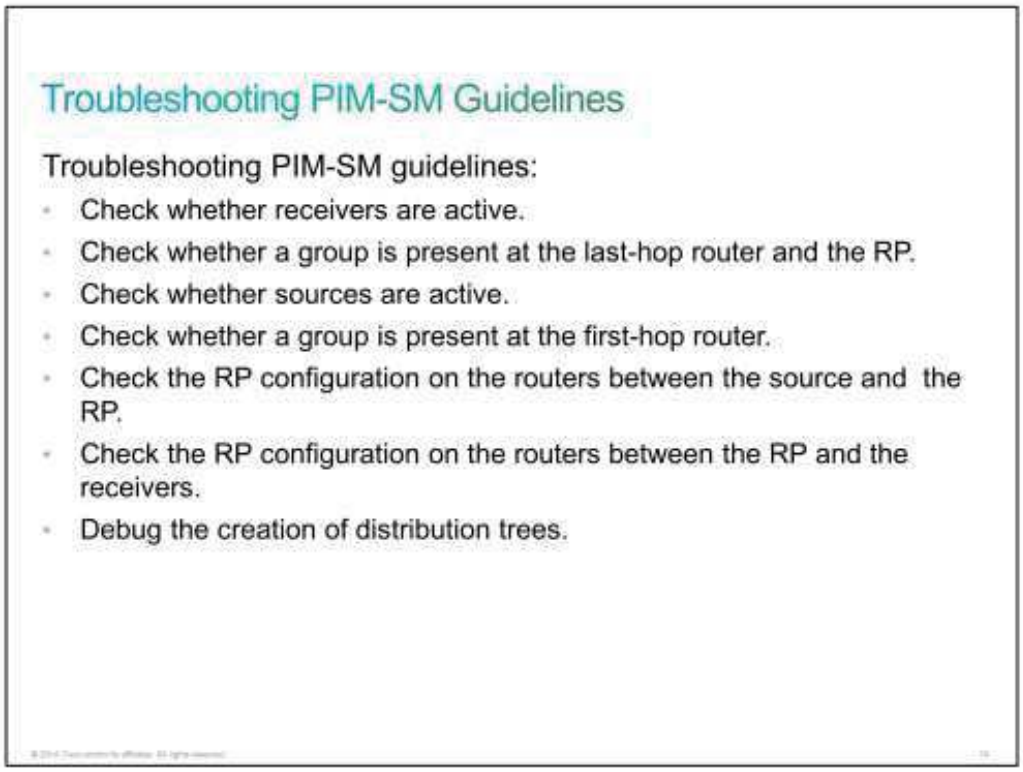
The Cisco IOS and IOS XE **show ip mroute** or the Cisco IOS XR **show mrib route** command is the most useful command in determining the state of multicast sources and groups. The output of the command generally represents a part of the multicast distribution tree with an incoming interface and a list of outgoing interfaces. The Cisco IOS, IOS XE, and IOS XR **summary** command option displays a one-line, abbreviated summary of each entry in the IP multicast routing table.



https://t.me/learningnets

# Troubleshooting PIM-SM Guidelines

When troubleshooting PIM sparse mode, start with the receiver segment and verify whether receivers are present and which groups they are reporting. Determine whether the groups are present on each router towards to the RP, on the RP, and on the routers between the RP and the last-hop router.




If receivers are recognized and groups are present, check to see whether sources are active and determine which groups are receiving which multicast traffic.

Verify whether the first-hop router recognizes the source traffic and has a group present.

Verify that all routers have PIM-SM configured on their interfaces, and turn on multicast distribution tree debugging.



https://t.me/learningnets

# Summary

This topic summarizes the key points that were discussed in this lesson.

## Summary

- PIM-SM operates with shared tree.
- DR joins the shared tree at RP when a receiver joins a group.
- First hop router builds source tree towards RP.
- Multicast tree can bypass the RP in certain situations.
- PIM uses control packets to send control-type information.
- PIM state is represented by entries in the multicast routing table.
- State is maintained by periodic updates.
- CLI commands show the contents of the multicast routing table.
- Different rules govern creation, deletion and maintenance of:
  - (*,G) states
  - (S,G) states
  - OIL
- Each state is described by several different flags.

## Summary (Cont.)

- PIM-SM control messages are passed between multicast routers.
- In the process of creating full multicast distribution tree:
  - the receiver can join first.
  - the source can join first.
  - receivers can be between first hop router and the RP.
- SPT switchover is triggered by exceeding the threshold value.
- When PIM-SM shared tree is pruned, interface is removed from all (*,G) and (S,G) entries.
- In case of SPT pruning only (S,G) entries are affected.
- The easiest way of configuring RP is to configure static RP; there is little redundancy in such setup.
- To successfully troubleshoot multicast implementation, check all components one by one.



https://t.me/learningnets

https://t.me/learningnets

# Lesson 2: Implementing PIM-SM Enhancements

## Overview

SSM is a datagram delivery model that best supports one-to-many applications. SSM uses all the benefits of sparse mode protocols, but eliminates the RPs and shared trees and only builds a SPT.

BIDIR-PIM is designed to be used for many-to-many applications. Multicast groups in bidirectional mode can scale to an arbitrary number of sources without incurring overhead due to the number of sources. BIDIR-PIM eliminates the registration and encapsulation process and the (S,G) state. Packets are natively forwarded from a source to the RP using the (*,G) state only. This capability ensures that only (*,G) entries will appear in multicast forwarding tables. It also ensures that the path taken by packets flowing from the participant to the RP and from the RP to the participant will be the same, by using bidirectional shared tree.

This lesson describes SSM, the variant of a basic PIM-SM model, and describes BIDIR-PIM, the derivative of a basic PIM-SM model.

Upon completing this lesson, you will be able to:

- Describe basic SSM concepts
- Describe an SSM scenario
- Describe IGMPv3 support for SSM
- Describe SSM mapping
- Describe SSM configuration
- Identify the need for bidirectional multicast
- Discuss multicast sources and receivers in a bidirectional multicast tree
- Discuss Bidirectional PIM traffic flow
- Discuss forwarding and tree building process
- Discuss the DF election process



https://t.me/learningnets

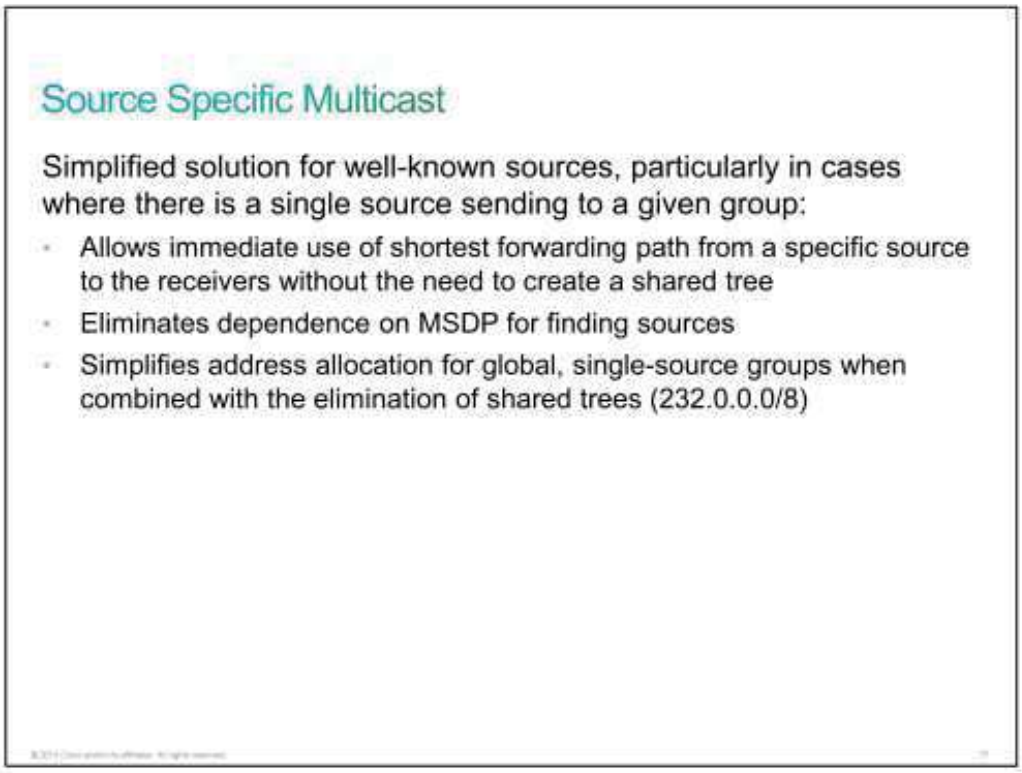
- Discuss the DF election messages
- Discuss the DF initial election
- Discuss winning the election as DF
- Discuss how the DF loses its path to the RP
- Discuss the situation when the DF dies or the metric changes
- Discuss Bidirectional PIM configuration

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# Source Specific Multicast

SSM is a new multicast concept and supports SSM applications. The SSM uses all the benefits of sparse mode protocols, but eliminates shared trees and only builds source-specific SPTs. These trees are built directly based on the receipt of group membership reports that request a given source. The SSM is described in RFC 3569.




SSM is suitable for use when there are well-known sources either within the local PIM domain or within another PIM domain. The MSDP, which is needed for interdomain multicast routing when regular PIM-SM is used within a domain, is no longer needed for SSM.

A dedicated multicast group address range of 232.0.0.0/8 is used exclusively for SPTs for SSM. Routers are prevented from building a shared tree for any of the groups from this address range. The address range 232.0.0.0/8 is assigned for global well-known sources.

SSM is a datagram delivery model that best supports one-to-many applications, also known as broadcast applications.



https://t.me/learningnets

## Source Specific Multicast (Cont.)

SSM for well-known sources characteristics:

- Allows last-hop router to send (S,G) join directly to source without the creation of a shared tree
- Allows first-hop router to respond to receiver-initiated join requests for specific sources within a group
- Supports elimination of shared-tree state in 232.0.0.0/8, simplifying address allocation

SSM allows the last-hop router to immediately send an (S,G) join toward the source. Thus, the PIM-SM (*,G) join toward the RP is eliminated, and the first-hop routers start forwarding the multicast traffic on the SPT from the very beginning. The SPT is built by receiving the first (S,G) join.

The assigned address range of 232.0.0.0/8 also simplifies address allocation problems because it is a global range for sources that must be well known. Implementations in routers must not build any shared tree for those groups.

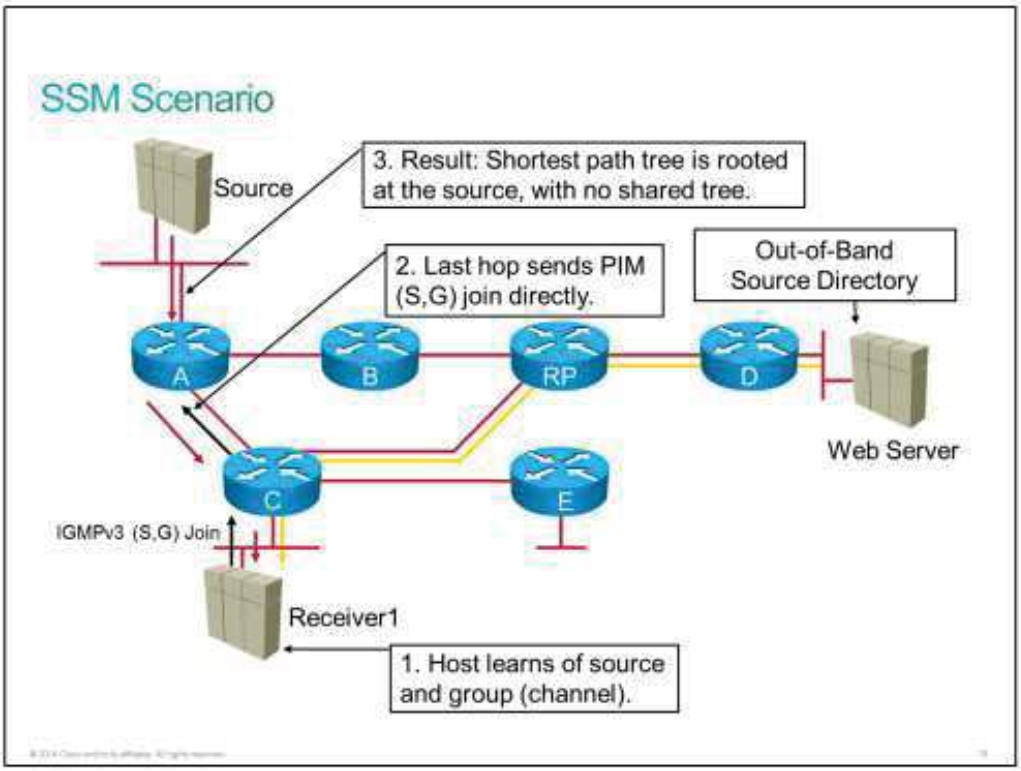
Source-specific groups may coexist with other groups in PIM-SM domains.



https://t.me/learningnets

# SSM Scenario

The prerequisite for SSM deployment is a mechanism that allows hosts to report not only the group that they want to join, but also the source for the group.




This mechanism is built into the IGMPv3 standard. With IGMPv3, last-hop routers may receive IGMP membership reports requesting a specific multicast source and group traffic flow. The router responds by simply creating an (S,G) state and triggering an (S,G) join toward the source.

Exactly how a host learns about the existence of sources may occur via directory service, session announcements directly from sources, or some out-of-band mechanisms (for example, web pages).

The result of building an SPT from the beginning is that all of the PIM-SM mechanisms that are associated with an RP are eliminated. RPs for SSM groups are not needed, because the discovery of sources is performed via some other method. In fact, routers must not build shared trees for groups in the SSM range (232.0.0.0/8).

There are several benefits of immediately building SPTs to a well-known source without the need for first building a shared tree. One of the main benefits concerns address management. Traditionally, it was necessary to acquire a unique IP multicast group address to ensure that a content source (such as a streaming video broadcast of an event) would not conflict with other possible sources sending on the shared tree.

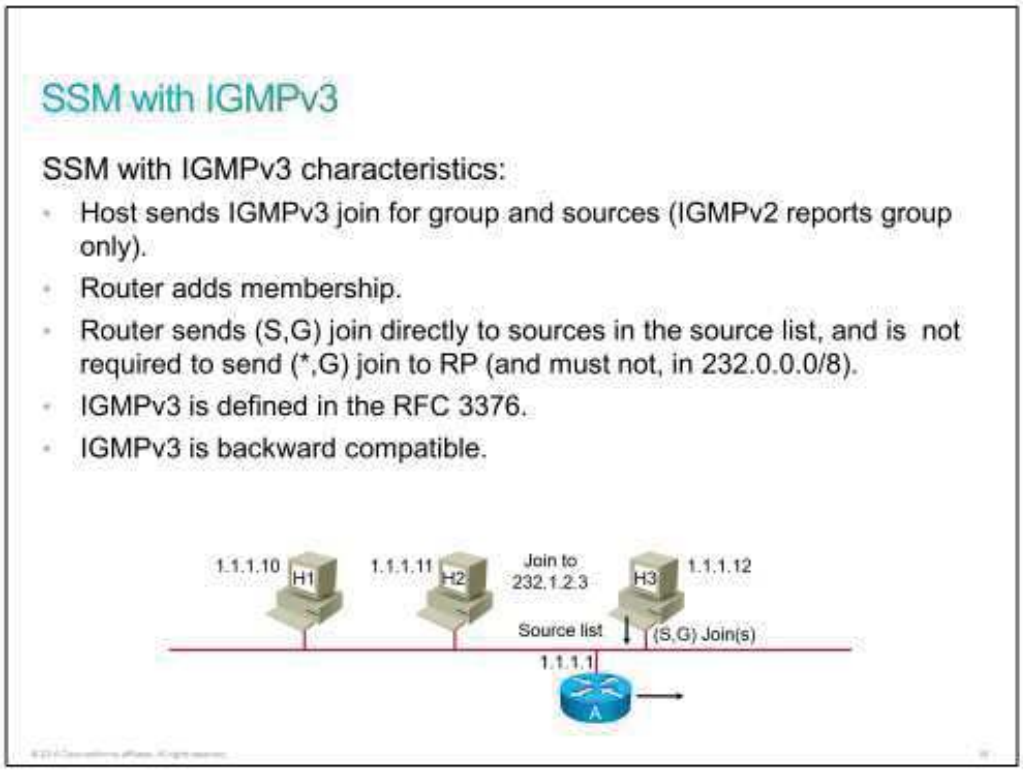
In SSM, a unique IP multicast group address is no longer necessary. Traffic from each source is uniquely forwarded using only an SPT. Thus, different sources may use the same SSM multicast group addresses without concern about intermixing traffic flows.



https://t.me/learningnets

# SSM with IGMPv3

The prerequisite for SSM deployment is IGMPv3, which allows a host to not only specify a group but also the source.




The specification of sources is accomplished via inclusion and exclusion lists (source list). When a router receives the IGMP report, it adds the group membership into its group table and initiates an (S,G) join directly to the source listed in a source list. The router must not send any (*,G) joins to the RP for groups in the 232.0.0.0/8 range.

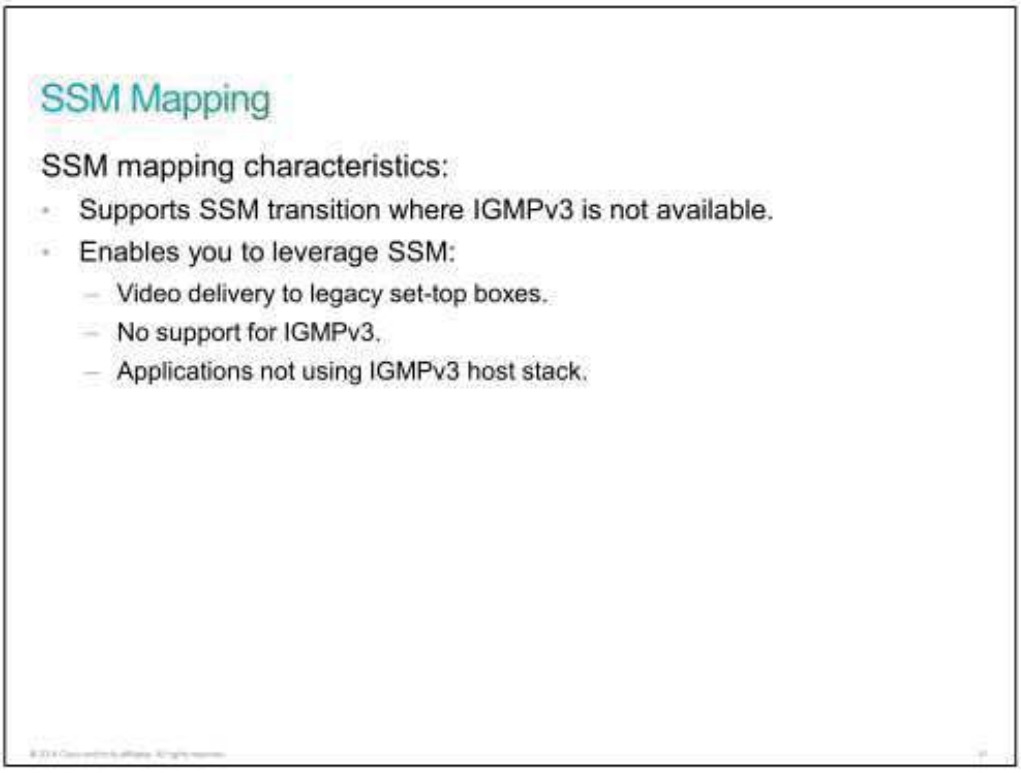
An SSM router must ignore all non-source-specific membership reports (such as IGMPv2 host reports) for groups in the SSM range.



https://t.me/learningnets

# SSM Mapping

SSM mapping supports SSM transition in cases where IGMPv3 is not available, or when supporting SSM on the end system is impossible or unwanted because of administrative or technical reasons.




SSM mapping enables you to leverage SSM for video delivery to existing set-top boxes that do not support IGMPv3. The mapping can also be used for applications that do not take advantage of the IGMPv3 host stack.

SSM mapping introduces a means for the last-hop router to discover the sources sending to the groups. When SSM mapping is configured, if a router receives an IGMPv1 or IGMPv2 membership report for a particular group, the router translates this report into one or more (S,G) channel memberships for the well-known sources associated with this group. SSM mapping enables the last hop router to determine the source addresses either by a statically configured table on the router or by consulting a DNS server. When the statically configured table is changed, or when the DNS mapping changes, the router will leave the current sources associated with the joined groups.

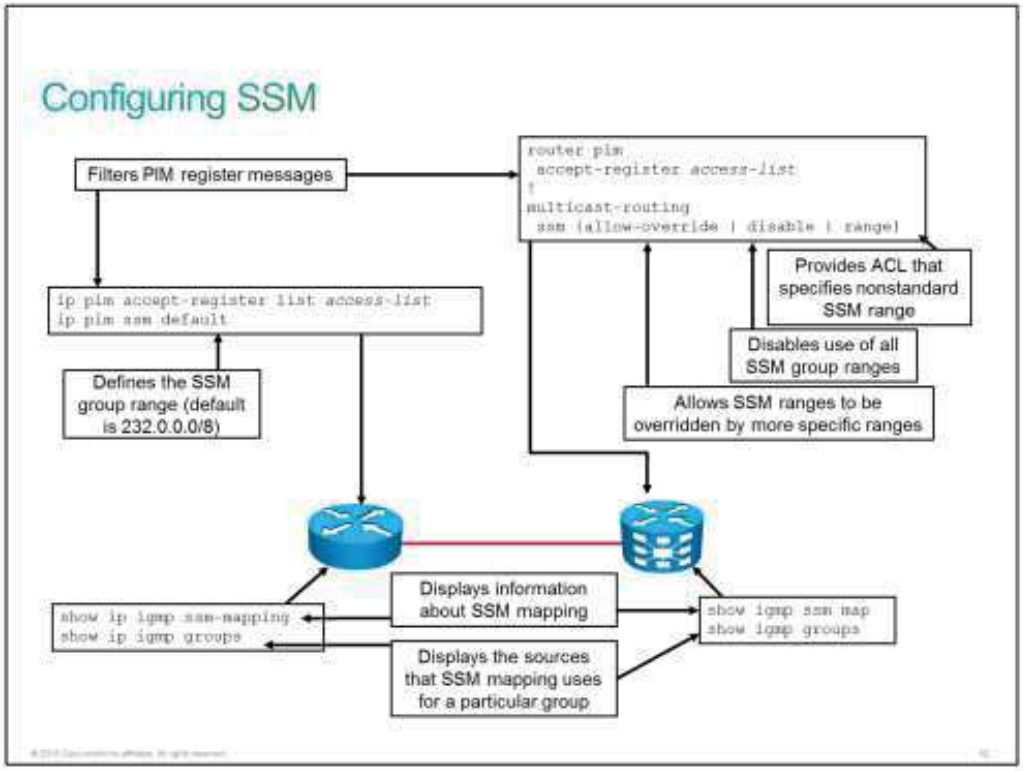
SSM mapping only needs to be configured on the last-hop router connected to receivers. No support is needed on any other routers in the network.



https://t.me/learningnets

# Configuring SSM

To configure a candidate-RP router to filter PIM register messages, use the Cisco IOS/IOS XE **ip pim accept-register** command or the Cisco IOS XR **accept-register** command. You can also use this command to prevent unauthorized sources in the SSM range from registering with the RP. If an unauthorized source sends a message to the RP, the RP will immediately send back a register-stop message.




The access list that is provided for the Cisco IOS/IOS XE **ip pim accept-register** or Cisco IOS XR **accept-register** command should only filter on IP source addresses and IP destination addresses.

Use the Cisco IOS/IOS XE **ip pim ssm default | range** command to define the SSM range of IP multicast addresses. The default range is 232.0.0.0/8. When an SSM range of IP multicast addresses is defined, no MSDP SA messages will be accepted or originated in the SSM range.

This Cisco IOS/IOS XE example shows how to configure SSM for the IP addressrange defined by access list 4:

```
access-list 4 permit 224.1.1.1
!
ip pim ssm range 4
```

To monitor IGMP and SSM, use these commands:

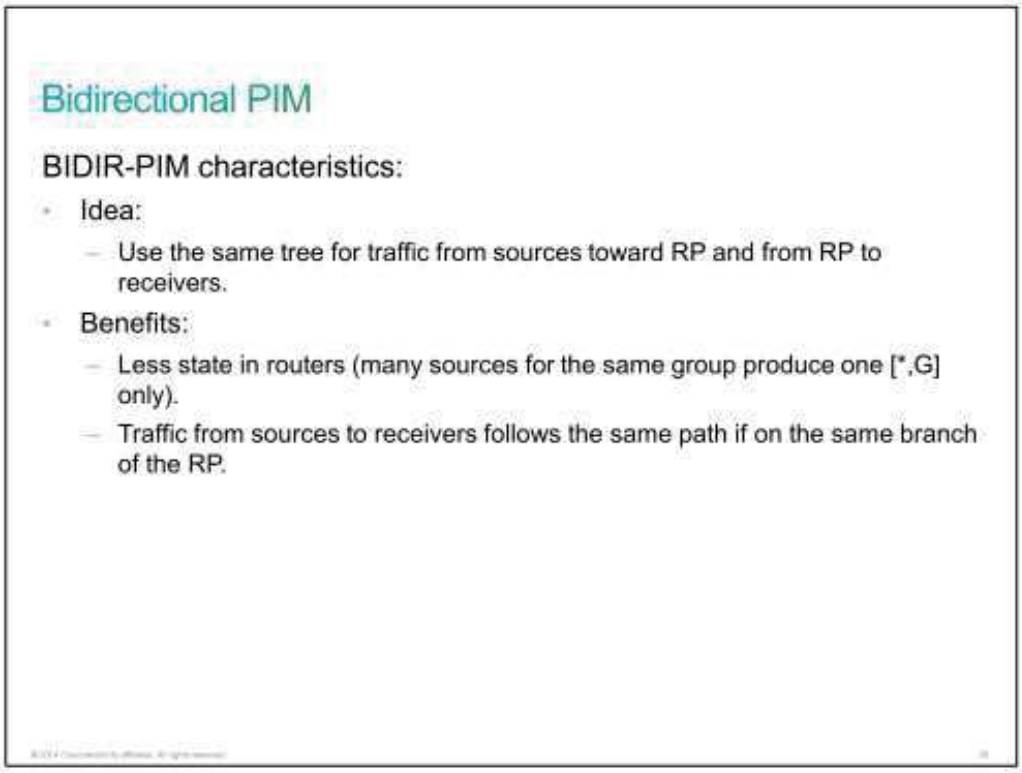
- The Cisco IOS/IOS XE **show ip igmp ssm-mapping** command or the Cisco IOS XR **show igmp ssm map** command displays information about SSM mapping.
- The Cisco IOS/IOS XE **show ip igmp groups** command or the Cisco IOS XR **show igmp groups** command displays the multicast groups with receivers that are directly connected to the router and that were learned through IGMP.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# Bidirectional PIM

PIM-SM is unidirectional in its native form. The traffic from sources to the RP initially flows encapsulated in register messages. This activity presents a significant burden because of the encapsulation and de-encapsulation mechanisms. Additionally, an SPT is built between the RP and the source, which results in (S,G) entries being created between the RP and the source.




Several multicast applications use a many-to-many multicast model where each participant is a receiver as well as a sender. The (*,G) and (S,G) entries appear at points along the path from participants and the associated RP. Additional entries in the multicast routing table increase memory and CPU utilization. An increase of the overhead may become a significant issue in networks where the number of participants in the multicast group grows quite large.

A good example is stock-trading applications where thousands of stock market traders perform trades via a multicast group. BIDIR-PIM eliminates the registration/encapsulation process and the (S,G) state. Packets are natively forwarded from a source to the RP using the (*,G) state only. This capability ensures that only (*,G) entries appear in multicast forwarding tables. The path that is taken by packets flowing from the participant (source or receiver) to the RP and from the RP to the participant will be the same by using a bidirectional shared tree.

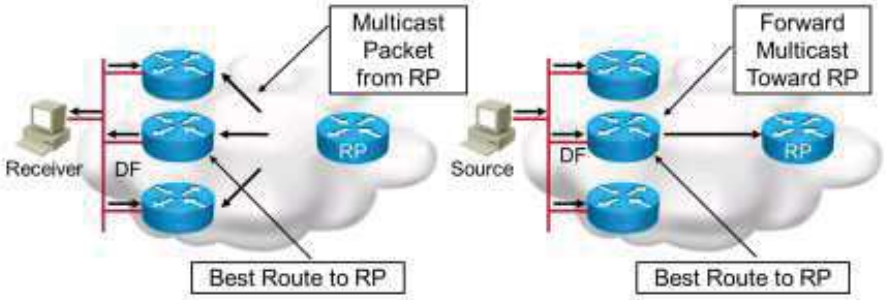


https://t.me/learningnets

## Bidirectional PIM (Cont.)

PIM modifications for bidirectional operation:

- On each link, the router with the best path to the RP is elected to be the DF.
- The DF is responsible for forwarding upstream toward the RP, as well as for downstream from the RP.
- No registration needed for local sources like in PIM-SM.




In BIDIR-PIM, the packet forwarding rules have been improved over PIM SM, allowing traffic to be passed up the shared tree toward the RP. To avoid multicast packet looping, BIDIR-PIM introduces a new mechanism called the DF which establishes a loop-free SPT rooted at the RP. The DF assumes the role of a designated router and has the following responsibilities:

- It is the only router that forwards packets traveling downstream (toward receiver segments) onto the link.
- It is the only router that picks up upstream-traveling packets (away from the source) off the link and forwards them toward the RP.

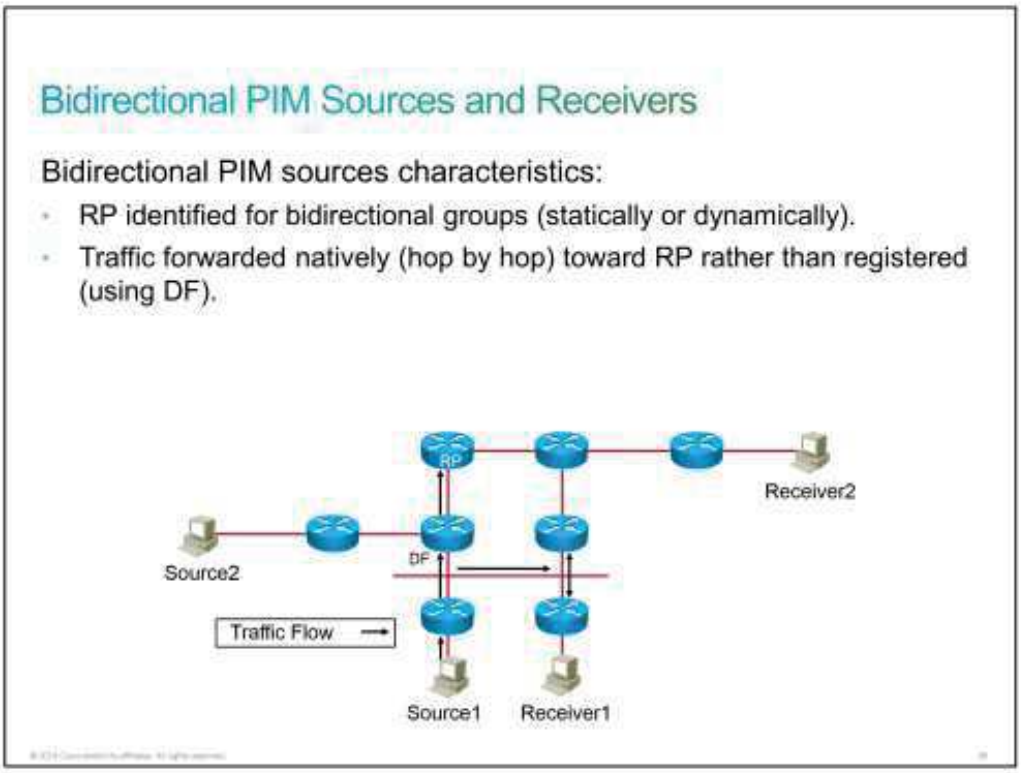
On every link in the network, the BIDIR-PIM routers participate in a procedure called DF election. The procedure selects one router as the DF for every RP of bidirectional groups. The router with the best unicast route to the RP is elected as a DF. There is also an election tie-breaking process if there are parallel equal cost paths to the RP. If the elected DF fails, it is detected via the normal PIM hello mechanism, and a new DF election process will be initiated.

 © 2014 Cisco Systems, Inc.

https://t.me/learningnets

# Bidirectional PIM Sources and Receivers

Initially, the routers responsible for sending (*,G) joins toward the RP and the routers responsible for forwarding group traffic toward the RP have to identify the group as bidirectional. This identification may be accomplished statically or dynamically via Auto-RP or the BSR mechanism. In PIMv2, routers discover that a group operates in bidirectional mode from the encoded-group address fields in PIM bootstrap and candidate-RP advertisement messages. The encoded-group address field has the bidirectional bit set. Regular PIM-SM groups may coexist with bidirectional groups.




The bidirectional PIM mechanism is based on the concept of a DF. A single DF for a particular BIDIR-PIM group exists on every link within a PIM domain. Elected on both multiaccess and point-to-point links, the DF is the router on the link with the best unicast route to the RP. A DF for a given RP is in charge of forwarding the following:

- Downstream traffic that flows down the shared tree onto the link.
- Upstream traffic that flows from the link toward the RP.

The DF does this for all the bidirectional groups that are served by the RP. The election mechanism for DF must ensure that all the routers on the link have a consistent view of the path toward the RP. Because the DF is associated with the RP, the term RP DF is very often used.

With BIDIR-PIM, as shown in the figure, traffic from Source1 is forwarded toward the RP by the DF. The same traffic from Source1 can also be forwarded directly on the branch toward the interesting receivers (Receiver1 in this example) without first reaching the RP.

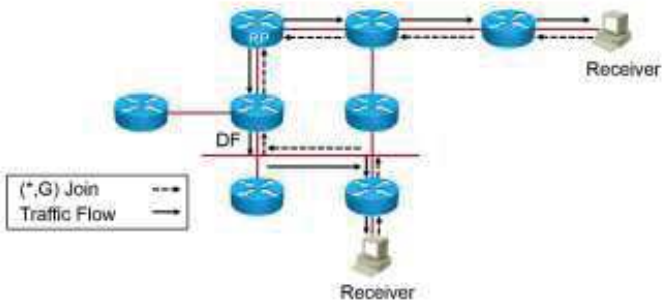


https://t.me/learningnets

## Bidirectional PIM Sources and Receivers

Bidirectional PIM receivers characteristics:

- Receivers join toward RP.
- Forwarding state is created with RPF toward RP as the IIF.




From the perspective of the receiver, there are no significant changes that are compared to the regular PIM-SM. The last-hop designated routers will forward (*,G) joins toward the RP serving the group. The only difference is that the DF performs the tasks of the designated router.

When a router receives a join message for a bidirectional group G, it must determine whether it is the DF on the link for this group. The router either inspects the (*,G) state or the RP DF election information when there is no (*,G) entry. If the router is the DF for the group, it follows the standard procedure for (*,G) joins. If the router is not the DF for the group, it ignores the received join.

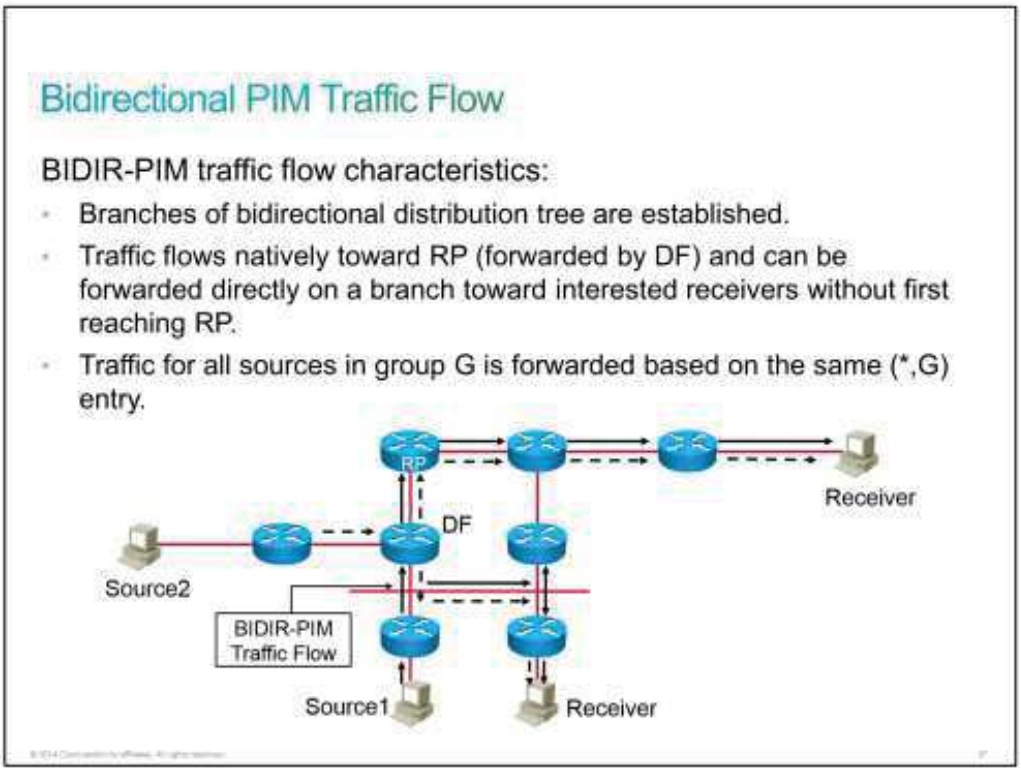
The shared tree is established between the receiver segments and the RP. The RPF interface or IIF is pointing toward the RP.

 © 2014 Cisco Systems, Inc.

https://t.me/learningnets

# Bidirectional PIM Traffic Flow

By forwarding (*,G) joins toward the RP, the network establishes branches of a shared tree. When sources start sending, their traffic is forwarded natively hop by hop toward the RP and further down established shared trees.




If the shared tree between the source and the RP exists, the entries in forwarding tables are used. If the shared tree does not exist, (*,G) entries are created and the election information for a DF is used to forward traffic via RPF interfaces toward the RP. Regardless of how many sources send the traffic for a certain group, only one (*,G) entry is created and used for this group.

The BIDIR-PIM router must forward the packet if either of these conditions exists:

- The packet is received on the IIF of the entry. This activity will prompt the router to forward downstream-traveling packets. This is the same procedure that exists with the regular PIM-SM model.
- The router is the DF for the respective RP and group for the interface on which the packet was received (only the DF forwards upstream).

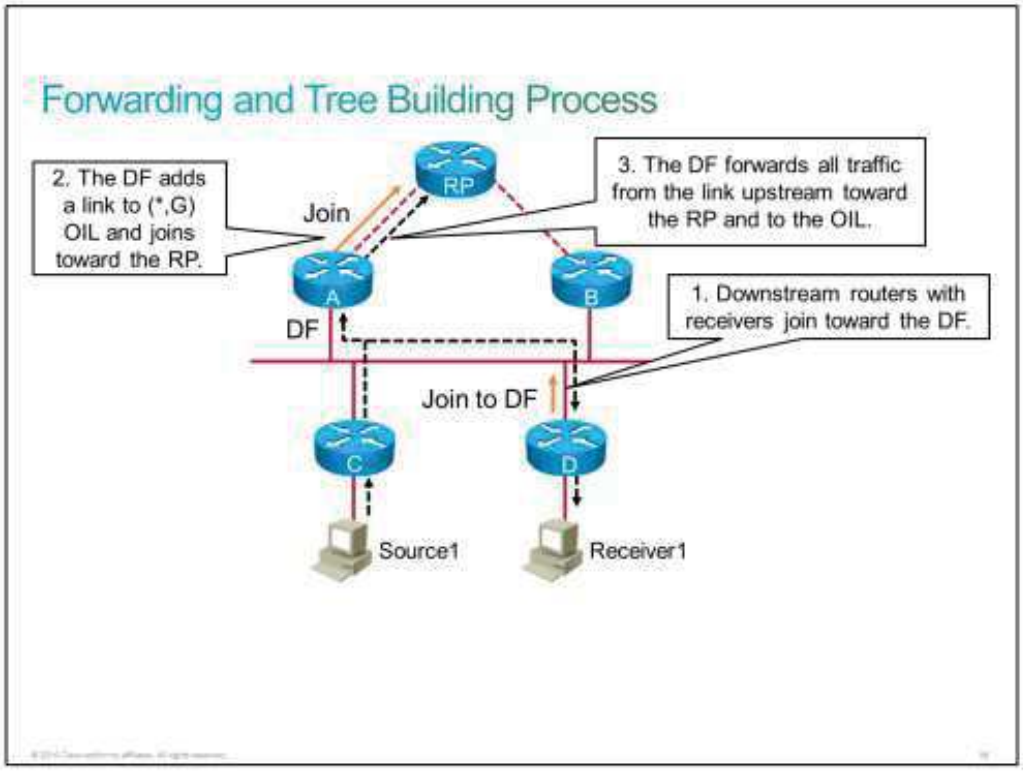
When the packet is forwarded, it is forwarded on all the interfaces in the OIL. Additionally, the packet is forwarded on the IIF toward the RP, excluding the interface on which the packet was received.



https://t.me/learningnets

# Forwarding and Tree Building Process

The forwarding and tree building process is described below:




Forwarding and Tree Building Process:

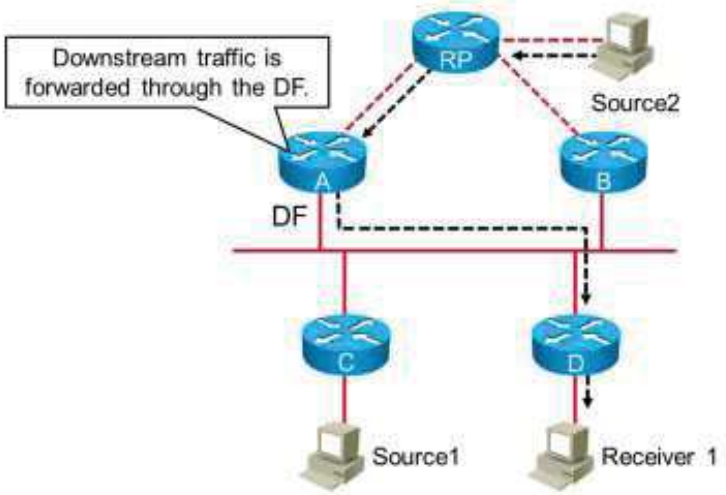
1. The DF is responsible for sending (*,G) joins toward the RP for the active bidirectional group. Downstream routers address their (*,G) joins to upstream DFs. This is accomplished by putting the IP address of the upstream DF in the upstream router field of a PIM join message.
2. When the DF receives a (*,G) join, it adds the link to the OIL of the (*,G) entry and joins toward the RP. If the interface exists in the OIL, the interface timer is refreshed.
3. The DF has the responsibility of forwarding multicast traffic in BIDIR-PIM. When multicast traffic is received on a link for which the router is the DF, the router must forward that traffic via its RPF interface toward the RP. Furthermore, the DF must forward the received traffic out all other interfaces in the (*,G) OIL, excluding the interface on which the traffic was received.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

## Forwarding and Tree Building Process (Cont.)

### Forwarding and Tree Building—Second Source

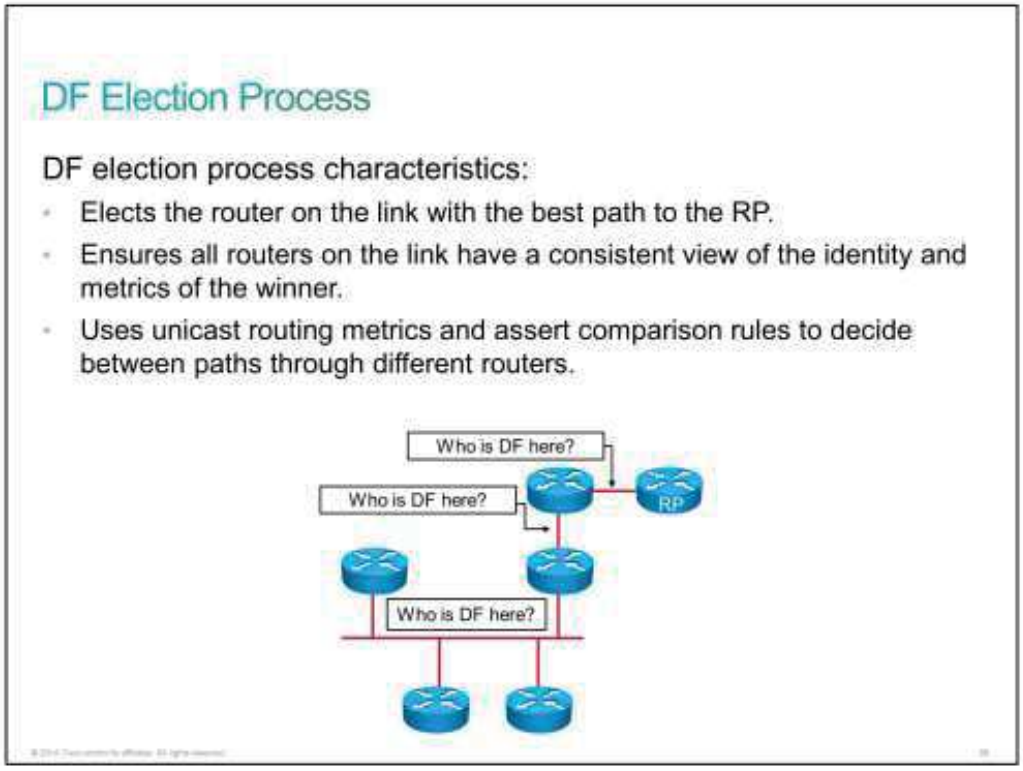





https://t.me/learningnets

# DF Election Process

In multiaccess networks, there may be parallel paths, which can lead to group members receiving duplicate packets from multiple routers. To avoid this problem, PIM uses assert messages to elect a single PIM forwarder to forward the traffic.




The BIDIR-PIM process of electing a DF on each link is similar to the PIM assert process.

The BIDIR-PIM DF mechanism ensures that all the routers on the link have a consistent view of the same RP. To perform the election of the DF for a particular RP, routers on a link need to exchange their unicast routing metric information for reaching the RP. The election of a DF is dependent on the RP, not an individual group.

The election process happens once only, when information on a new RP becomes available. However, an update to the election is needed under the following conditions:

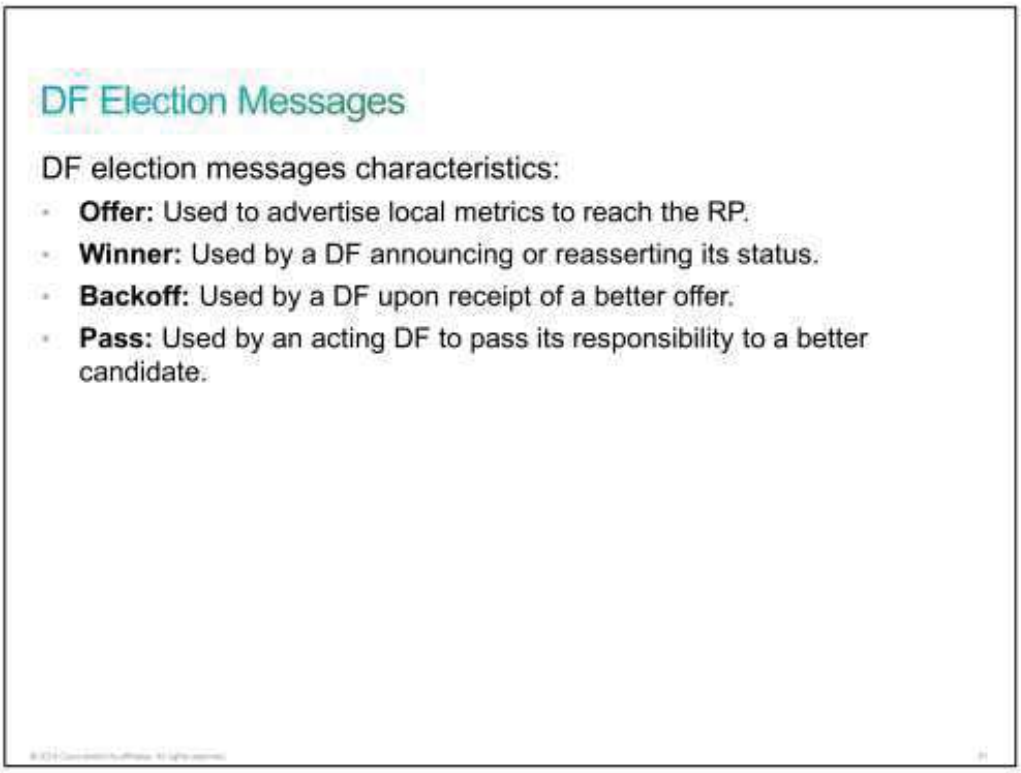
- A change occurs in the unicast metric to reach the RP for any of the routers on the link.
- The interface on which the RP is reachable changes to an interface for which the router was previously the DF.
- A new PIM neighbor is established on a link.
- The elected DF dies.



https://t.me/learningnets

# DF Election Messages

The DF election messages are described below.




The DF election mechanism is based on four control messages that are exchanged between the routers on the link:

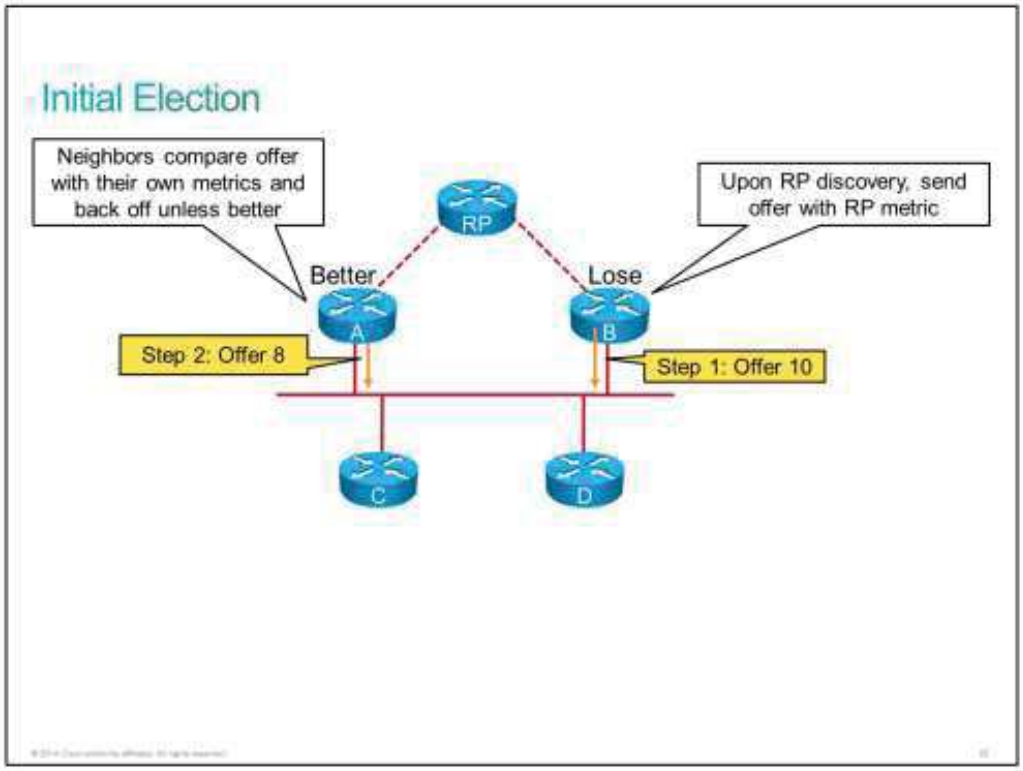
- **Offer message:** This message is used to advertise a router unicast metric to reach the RP. The metric is compared by the other routers participating in the election of a DF with their metric to reach the RP.
- **Winner message:** This message allows the winning router to announce to all routers on the link that it has won the DF election process. The DF also sends this message to reassert its status as the elected DF.
- **Backoff message:** This message is sent by the currently elected DF when it receives an offer message containing a better metric to the RP than its own. The DF records the received information and responds with a backoff message. The backoff message, therefore, is used to acknowledge that the sender (the active DF) has a worse metric back to the RP than the offering router. When the offering router receives the backoff message, it will assume the duties as the newly elected DF. The newly elected DF sends a winner message after a short period of time, while it allows unicast routing to stabilize.
- **Pass message:** This message is used by the acting DF to pass its function to another router that is offering a better metric. The old DF stops its tasks when the transmission is made.



https://t.me/learningnets

# Initial Election

When a router discovers that a new RP exists but a DF does not exist yet, it sends an offer message. The message contains the router metric to reach the RP as well as the identity of the router. The offer message is periodically retransmitted, and this period is called the offer interval.




When a router receives an offer message, it compares the offered metric with its own metric back to the RP. If the router learns about a better metric from a neighbor, it stops sending offer messages for an amount of time that is three times the offer interval. If, after this period, no winner is elected, the router will restart the election. A router takes up the function of the DF after sending three offers without receiving any offers from any other neighbor.

If the router receives an offer and it has a better metric back to the RP, it actively starts participating in the election by sending its own offer messages. The router with a better metric than the offered one sends its own offer. This offer includes its metric to the RP and its identity.

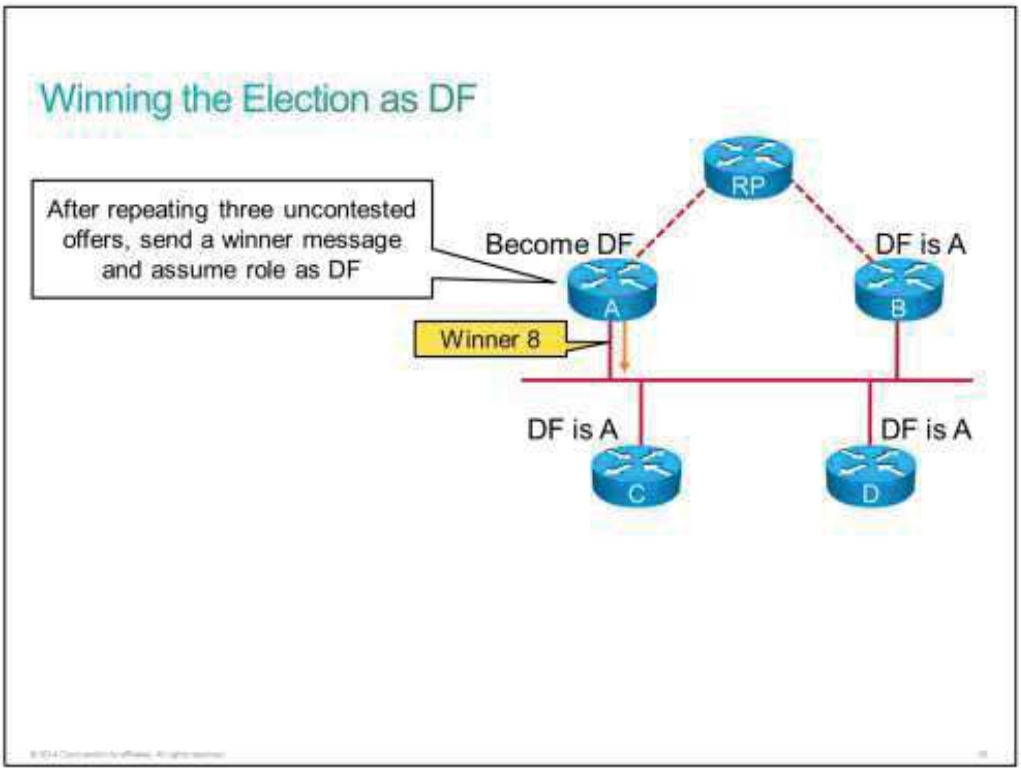
In the figure, router B receives an offer with a better metric, assumes that it has lost the election, and backs off. The backoff means that the router will not send any offers for at least three times the offer-interval period. The router will wait for a winner message to indicate the DF election process has completed. If, after the offer-interval period, a winner message is not received, the DF election process restarts.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# Winning the Election as DF

In the figure, router A has sent its offer three times and has not received another offer with a better metric. Winning the election as DF, router A assumes the role and announces its election by transmitting a winner message on the link. It also transmits its identity (IP address) and the metric it is using.



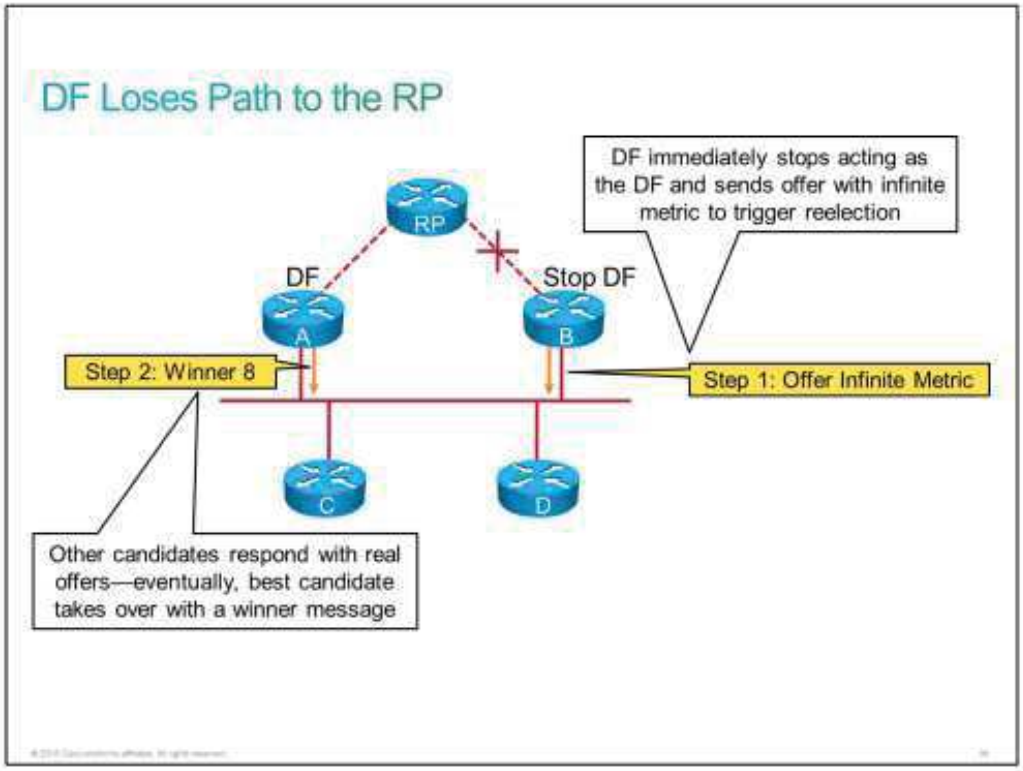

Routers that receive a winner message stop participating in the election and record the identity and metrics of the winner.



https://t.me/learningnets

# DF Loses Path to the RP

When the path to the RP currently used by the DF switches through the link for which it is the DF, it may no longer provide forwarding services. The DF immediately stops being the DF and restarts the election by sending an offer message with an infinite metric. If no better offer is received, an infinite offer is repeated periodically.



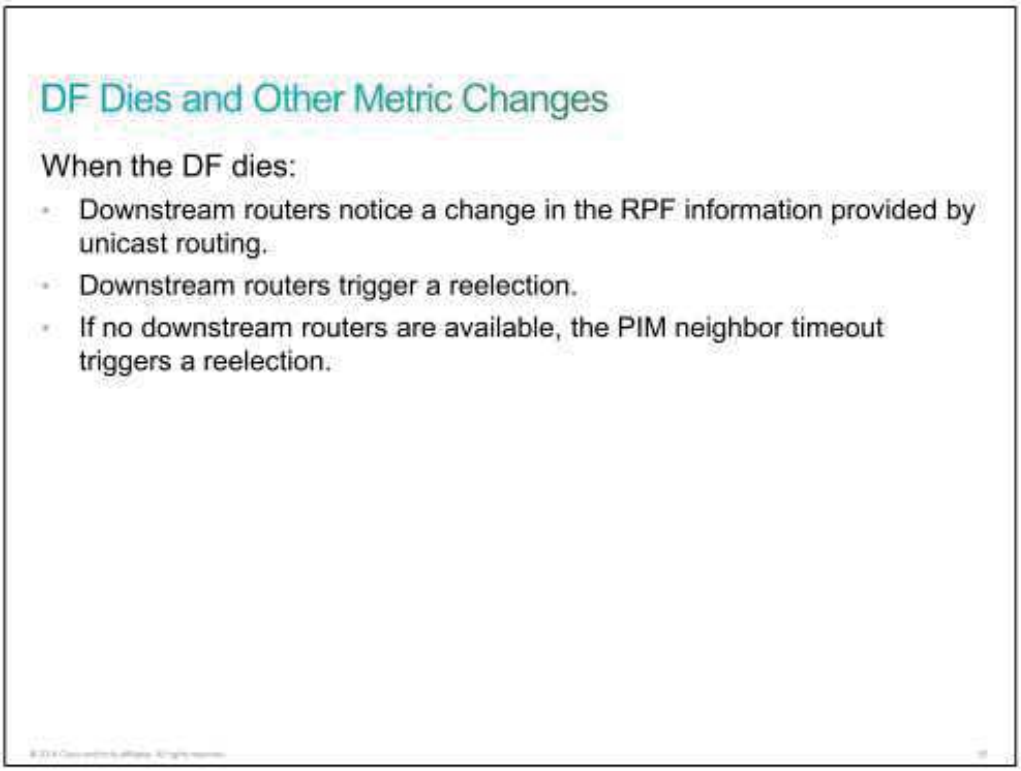

After the router acting as a DF loses the path to the RP and its RPF interface becomes the same as the interface for which it is the DF, the procedure is similar to the standard DF election procedure. Routers that receive an infinite offer respond with their offers. The router with the best offer assumes the DF function and transmits a winner message.



https://t.me/learningnets

# DF Dies and Other Metric Changes

The speed at which a new DF is elected after the original DF goes down depends on whether there are any downstream routers on the link.




For downstream routers, the RPF neighbor (that is also the DF) will change, and the routers will initiate the reelection process by sending offer messages. If the RP is reachable through the link via another upstream router, the downstream router will use an infinite metric in its offer message.

If no downstream routers are available, the only way for other upstream routers to detect the failure of a DF is by the timeout of the PIM neighbor information. This process takes longer.



https://t.me/learningnets

## DF Dies and Other Metric Changes (Cont.)

Other metric changes characteristics:

- When the RP metric at a non-DF router changes to a value that is worse than that of the acting DF, then no action is taken.
- When the metric at the DF improves, a winner message may be sent to update information in neighboring routers.
- When the metric at the DF becomes worse, three winner messages are sent to give a better candidate the opportunity to respond with an offer.

There are some other situations where the metric to the RP changes. When the metric of a router that is not the DF changes to a value that is worse than that of the current DF, no action is taken.

There may also be changes to the metric of the current DF. If the DF metric becomes worse than before (assuming that the DF still has a path to the RP), the DF sends a set of three randomly spaced winner messages with the new metric. Routers that receive this message and have a better metric may respond with an offer message. This action triggers the same process that occurs when the metric of a router that is not the DF becomes better than the metric of a router that is the current DF. All routers assume that the DF has not changed until they see a pass or winner message indicating the change.

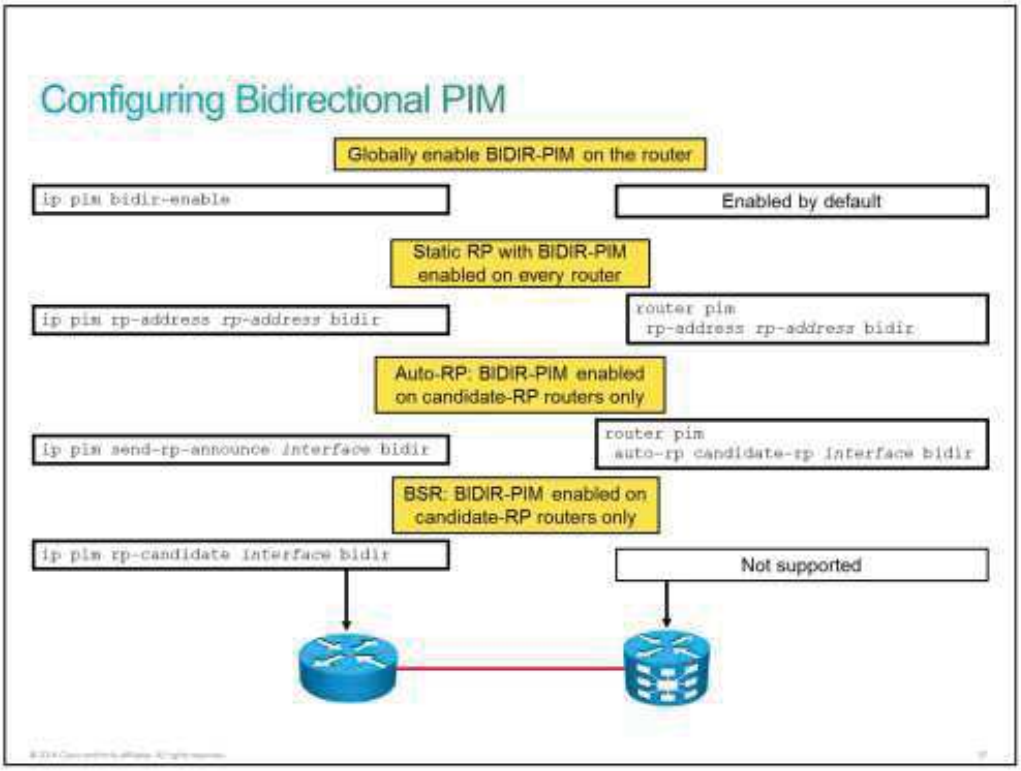
If the routing metric at the DF changes to a better value, a single winner message is sent advertising the new metric.



https://t.me/learningnets

# Configuring Bidirectional PIM

Bidirectional PIM configuration is described below:




The Cisco IOS/IOS XE **ip pim bidir-enable** command enables BIDIR-PIM on the router. The Cisco IOS XR router has BIDIR-PIM enabled by default. A router with BIDIR-PIM enabled will send PIM hello messages with the bidirectional mode option. When BIDIR-PIM is enabled, bidirectional PIM can be configured with static RP or dynamic RP methods, as shown on the figure.

| | |
|---|---|
| **Note** | Auto-RP and BSR operations and configurations will be covered in more detail in a later lesson. |



https://t.me/learningnets

# Summary

This topic summarizes the key points that were discussed in this lesson.

## Summary

- SSM is a simplified solution for well-known sources
- Best case for SSM is a case where there is a single source sending to a given group
- Host sends IGMPv3 join for group and sources
- SSM mapping supports SSM transition where IGMPv3 is not available
- SSM is configured differently on IOS and IOS XR platforms
- Bidirectional multicast uses the same tree for traffic from sources toward RP and from RP to receivers

## Summary (Cont.)

- Source traffic is forwarded natively, while receivers still join towards the RP
- Traffic flows natively toward RP, forwarded by DF
- DF election elects the router on the link with the best path to the RP
- When the DF dies or metric changes, downstream routers notice a change in the RPF information provided by unicast routing
- Bidirectional PIM is enabled by default on IOS XR

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# Lesson 3: Implementing Interdomain IP Multicast

## Overview

Interdomain IP multicast forwarding is constrained by the lack of proper standardized protocols for building interdomain distribution trees. MSDP is a solution that allows interdomain multicast routing by announcing sources from other domains into local domains. This capability interconnects multicast distribution trees.

This lesson explains the basic concepts of MSDP and its use in the IP multicast environment. This lesson also explains the steps that are associated with configuring MSDP.

Upon completing this lesson, you will be able to:

- List the requirements for service providers to exchange IP multicast traffic
- Explain RFC 2770 GLOP addressing
- Discuss the role of SSM in interdomain IP multicast
- Discuss the role of MSDP in interdomain IP multicast
- Describe multicast service provider requirements
- Describe the MSDP protocol
- Describe basic MSDP concepts
- Describe the MSDP neighbor relationship
- Describe MSDP messages
- Discuss how SA messages are processed
- Discuss how SA messages are originated



https://t.me/learningnets

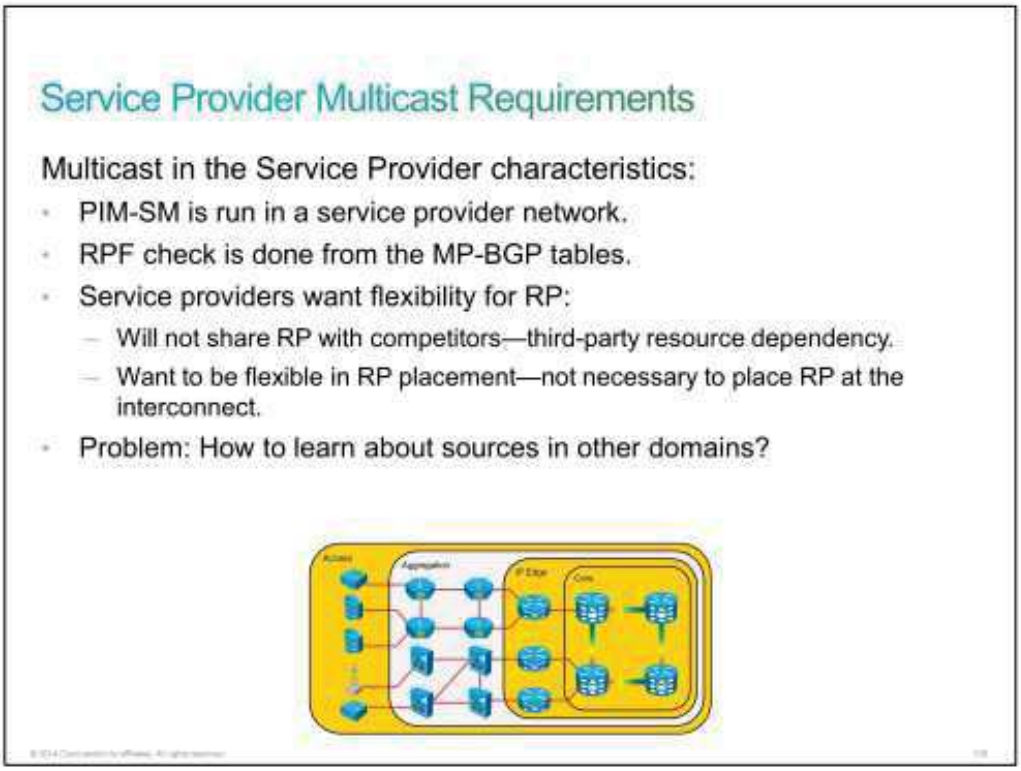
- Discuss MSDP MD5 password authentication
- Discuss MSDP configuration
- Discuss MSDP verification

 © 2014 Cisco Systems, Inc.

https://t.me/learningnets

# Service Provider Multicast Requirements

Modern service providers have some mature technologies at their disposal when they implement IP multicast. Within their network (domain), they mainly use PIM-SM. For interdomain multicast, service providers can use the MP-BGP to perform RPF checks for multicast traffic on interdomain links.




PIM-SM implementations require RP, and service providers want flexibility with respect to the following:

- **Ownership of the RP:** Service providers do not want to be dependent on RPs in other domains. If a set of customers for a service provider is communicating on multicast group G and the RP for that group exists in the domain of a competitor, a problem can occur if the RP goes down. The first service provider is dependent on its competitor, then, to fix the failed RP to enable its customers to continue to communicate.
- **Placement of the RP:** Service providers want to be able to place multiple RPs in their network at different locations. They do not want to be constrained by a single interconnect point.

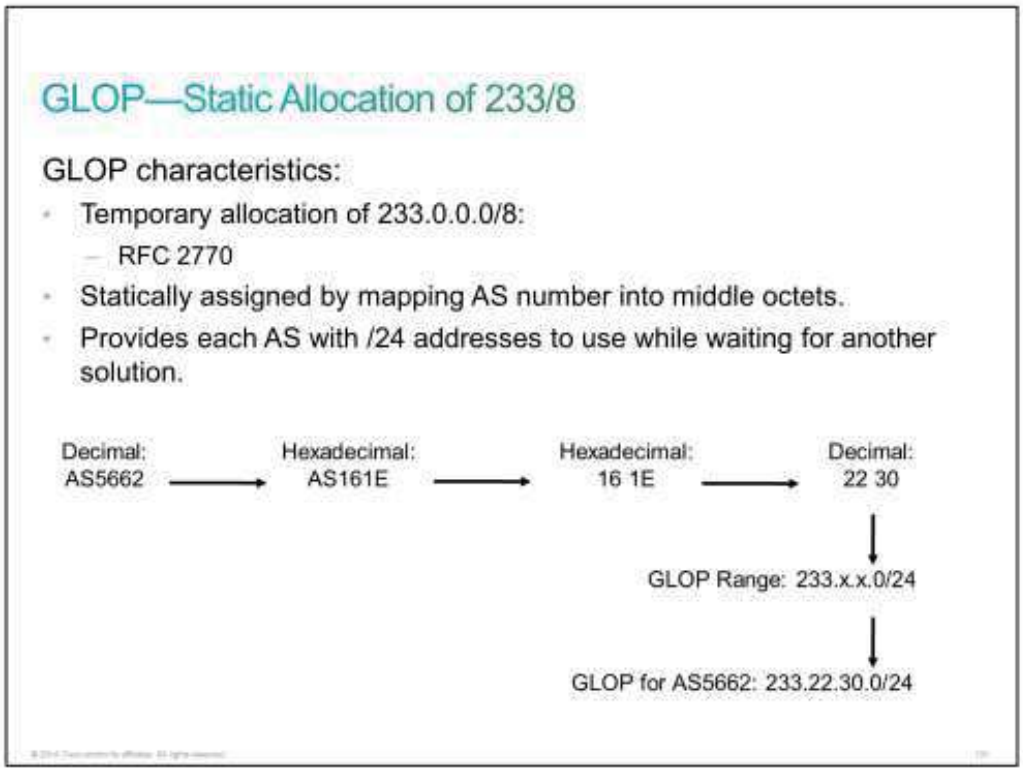
For service providers to maintain their own RPs for all their multicast groups, they need a mechanism that enables them to learn about active sources in other domains.



https://t.me/learningnets

# GLOP—Static Allocation of 233/8

RFC 2770 proposes that the 233.0.0.0/8 address range be reserved for statically defined addresses by organizations that already have an AS number reserved. This practice is called GLOP addressing. The AS number of the domain is embedded into the second and third octets of the 233.0.0.0/8 address range.



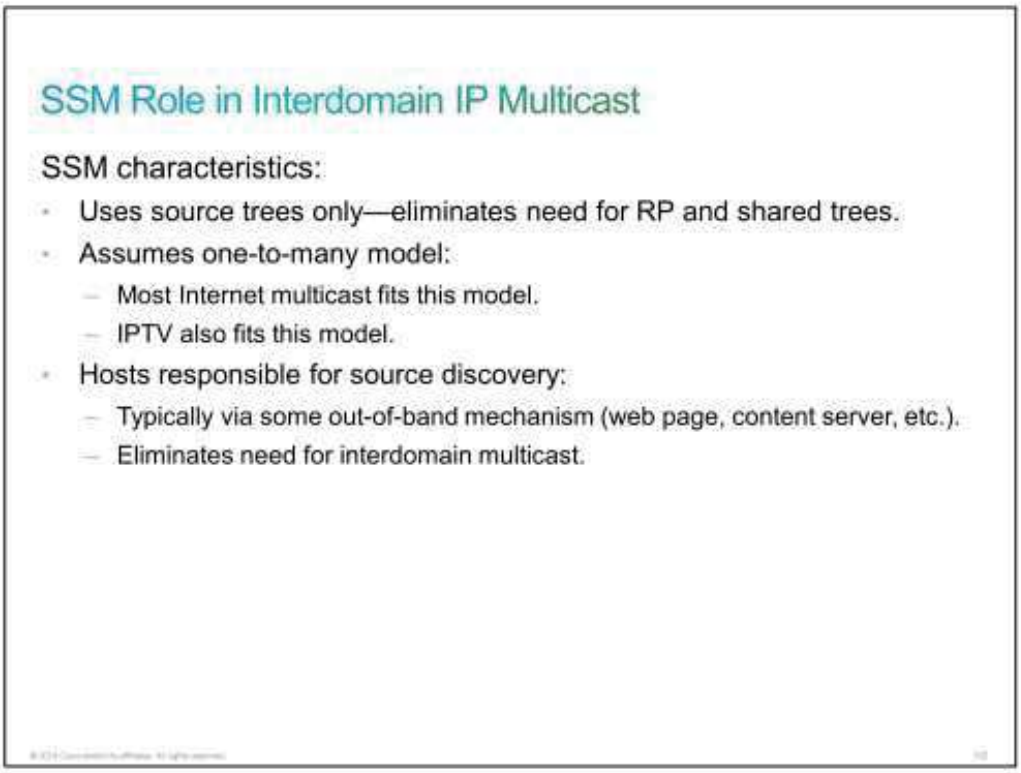

For example, the AS number 5662 is written in hexadecimal format as 161E. Separating the two octets (16 and 1E) results in 22 and 30 (in decimal format). These values result in a subnet of 233.22.30.0/24 that would be globally reserved for AS 5662 to use.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# SSM Role in Interdomain IP Multicast

An additional approach to interdomain multicast routing is SSM. SSM eliminates the need for shared trees for some well-known sources. A dedicated multicast group address range 232.0.0.0/8 is used exclusively for SPTs for SSM. Routers are prevented from building a shared tree for any of the groups from this address range. The address range 232.0.0.0/8 is assigned for global well-known sources. The last-hop routers send joins directly to the source as soon as they learn about multicast groups that want to receive traffic from those sources. SSM is suitable for well-known sources within a domain or in another domain.






https://t.me/learningnets

## SSM Role in Interdomain IP Multicast (Cont.)

- Hosts join a specific source within a group:
  - Content identified by specific (S,G) instead of (*,G).
  - Hosts responsible for learning (S,G) information.
- Last-hop router sends (S,G) join toward source:
  - Shared tree is never joined or used.
  - Eliminates possibility of content jammers.
  - Only specified (S,G) flow is delivered to host.
- Simplifies address allocation:
  - Dissimilar content sources can use same group without fear of interfering with one another.

It is a variant of a PIM-SM that supports SSM applications. SSM adopts all the benefits of sparse mode protocols but eliminates shared trees, only building source-specific SPTs. These trees are built directly upon receiving group membership reports that request a given source.

SSM is a datagram delivery model that best supports one-to-many applications, also known as broadcast applications. Interested receivers can learn the well-known sources with some type of out-of-band mechanism, such as web pages, content servers, and so on.

When regular PIM-SM is used within a domain, no other protocol for interdomain multicast routing is needed for SSM.



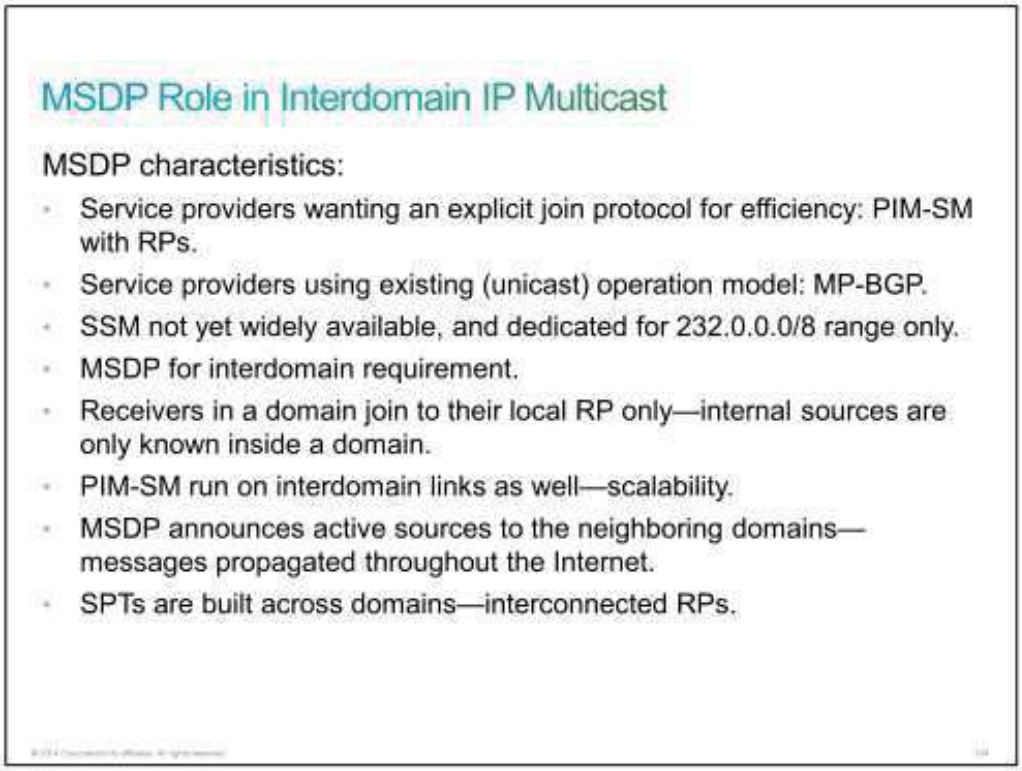
https://t.me/learningnets

# MSDP Role in Interdomain IP Multicast

Service providers want an explicit join model for their multicast deployments because this model is proven to be scalable. PIM-SM, with an RP as a root of a shared tree, is used inside a provider domain. Service providers want to maintain their own RP and do not want to rely on RPs in other domains.

Service providers also want to minimize the impact of multicast on their operations model. This can be achieved by using MP-BGP to carry both unicast and multicast routing information. Because MP-BGP uses the same configuration, maintenance, and policy controls as BGP, virtually no new training is required of operations staff.

SSM is a solution that eliminates the need for RPs. However, SSM is not yet widely implemented and is restricted to the 232.x.x.0/24 multicast address range.




MSDP turned out to be a simple and elegant solution to interconnect PIM-SM domains. It is also the easiest solution for interdomain multicast routing today. MSDP announces sources in one domain to other domains, and it allows RPs to build interdomain multicast distribution trees.

The idea builds on the active participation of the RP inside a domain and interconnection with RPs in other domains. The RP inside a domain is aware of active groups, since shared trees for these groups are rooted at the RP. The RP also knows active sources in a domain, because they register with the RP. The sources are only known within the domain.

To propagate information outside a domain, PIM-SM must first be implemented on interdomain links to provide for (S,G) join forwarding toward sources in other domains.

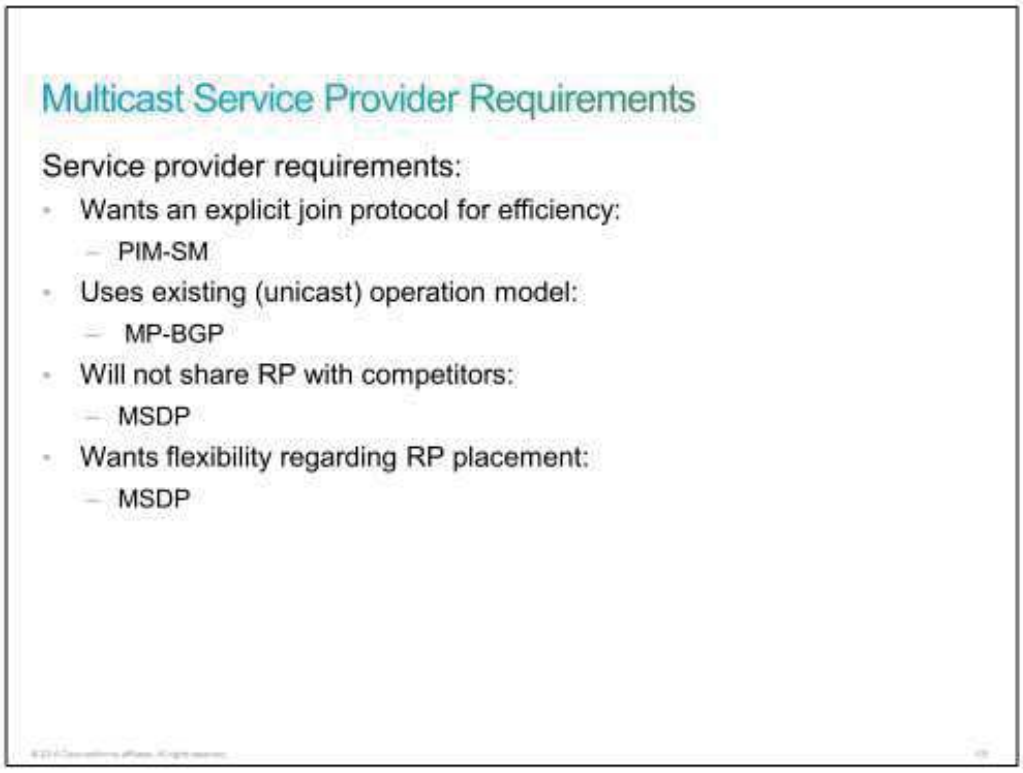
MSDP is a mechanism to connect multiple PIM-SM domains. MSDP allows multicast sources for a group to be known to all RPs in different domains. Each PIM-SM domain uses its own RPs and does not have to depend on RPs in other domains. An RP runs MSDP over TCP to discover multicast sources in other domains. Only SPTs are built between domains.



https://t.me/learningnets

# Multicast Service Provider Requirements

The service provider multicast requirements have all now been met using a combination of PIM-SM, MP-BGP, and MSDP.




If an explicit join protocol is required, PIM-SM can be used to meet the requirement.

If an existing unicast operations model is used, the extension of BGP to MP-BGP permits both unicast and multicast traffic flows to be configured and managed using the same set of tools.

If an RP does not need to be shared with competitors, MSDP permits each PIM-SM domain to have its own RP for each group.

If flexibility regarding placement of the RP is required, MSDP allows RPs to be placed anywhere in the PIM-SM domain as long as they are linked to other RPs in other domains.

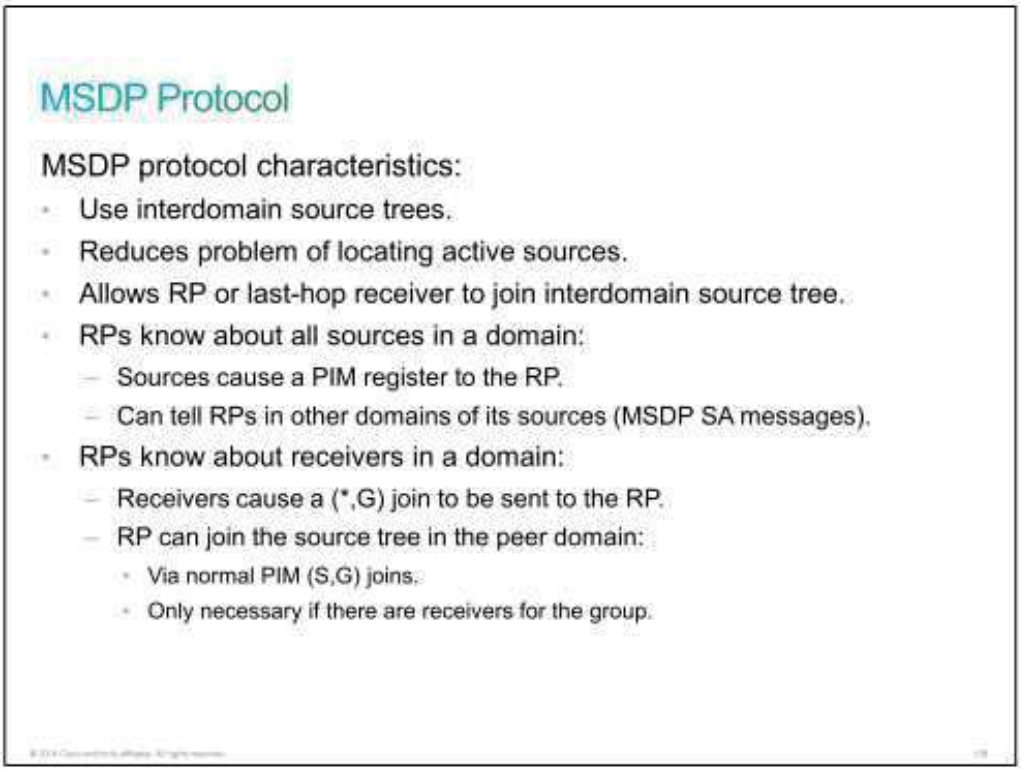
| | |
|---|---|
| **Note** | It is not necessary to have a full mesh of MSDP connections. It is sufficient to have a single MSDP connection that leads to the rest of the MSDP speakers (RPs) in the Internet. |



https://t.me/learningnets

# MSDP Protocol

By abandoning the notion of interdomain shared trees and using only interdomain source trees, the complexity of interconnecting PIM-SM domains is reduced considerably.




The remaining problem becomes one of communicating the existence of active sources between the RPs in different PIM-SM domains.

The RP can join the interdomain source tree for sources that are sending to groups for which the RP has receivers—group members. This is possible because the RP is the root of the shared tree, which has branches to all points in the domain where there are active receivers.

If the RP either has no shared tree for a particular group or a shared tree whose OIL is empty (null), it does not send a join to the source in another domain.

Once a last-hop router learns of a new source outside the PIM-SM domain (via the arrival of a multicast packet from the source down the shared tree), it can send a join toward the source. It can then join the source tree (SPT switchover mechanism).

The MSDP is based on a simple idea and works well in the interdomain multicast connections of today. The expected growth of the multicast-enabled Internet could also result in MSDP scalability problems. MSDP SA messages, used to tell RPs in other domains about multicast sources and groups, are sent every 60 seconds. If there is a large number of sources, this setup could become an issue.

The entire concept of MSDP depends on the RPs in the interconnected domains. The RPs in a domain must know about all the groups and all the sources in a domain. As a result, MSDP can only work with PIM-SM.



https://t.me/learningnets

When a source goes active in a PIM-SM domain, the first-hop router immediately informs the RP of this activation via PIM register messages. The (S,G) state for the source is kept alive in the RP by normal PIM-SM mechanisms, as long the source is actively sending. As a result, the RP can inform RPs in other PIM-SM domains of the existence of active sources in its local domain. MSDP SA messages accomplish this action.

Receivers (group members) cause last-hop routers to send (*,G) joins to the RP to build branches of a shared tree for a group. If an RP has a (*,G) state for a group, and the OIL of the (*,G) entry is not empty, the RP knows that it has active receivers for the group. Therefore, when it receives an SA message announcing an active source for group G in another domain, it can send (S,G) joins toward the source in the other domain.

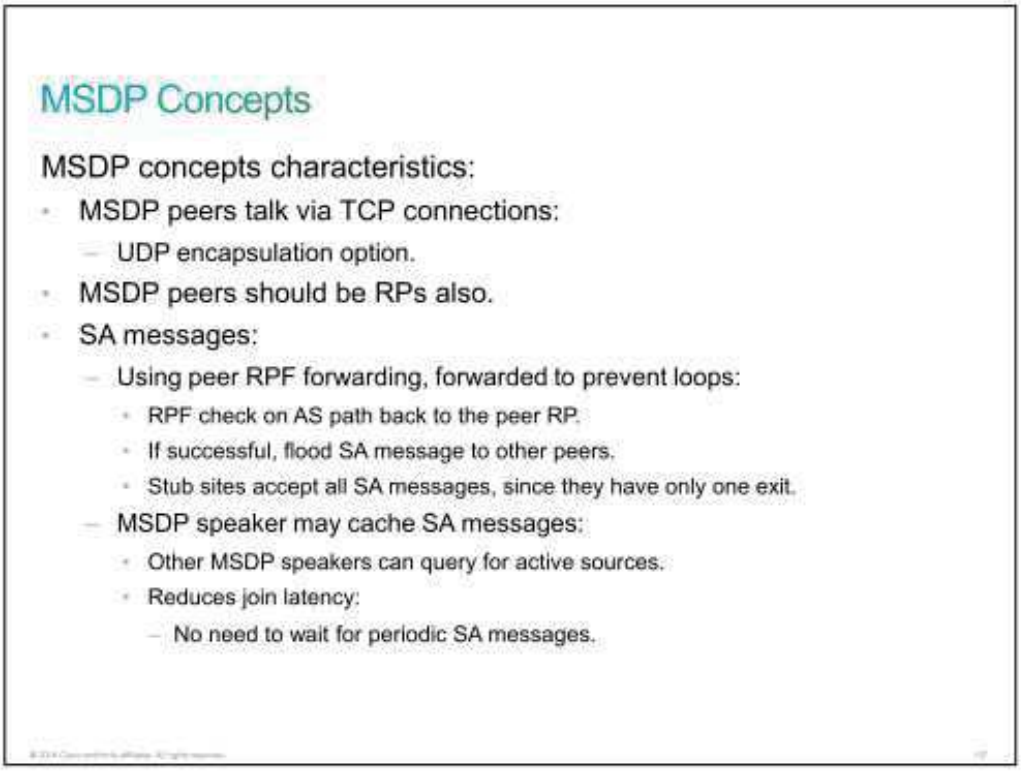
Anycast RP is a useful application of MSDP. Anycast RP is an intradomain feature that provides redundancy and load-sharing capabilities. With Anycast RP, a source may register with one RP and receivers may join a different RP, so a method is needed for RPs to exchange information about the active sources. This information exchange is done with MSDP. Anycast RP will be discussed in more detail in a later lesson.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# MSDP Concepts

Basic MSDP concepts are described below:




The MSDP peers should also be the RPs in respective domains. They are connected via MSDP sessions (TCP sessions, although UDP is an option).

SA messages are periodically (60 seconds) originated by the RPs of a domain for sources that are active in their local domain. These SA messages are sent to all active MSDP peers.

When an MSDP speaker receives an SA message from one of its peers, it performs an RPF check, then forwards (floods) it to all its other peers. A peer RPF check is performed on the arriving SA message (using the originating RP address in the SA message) to ensure that it was received via the correct AS path. The BGP routing table is examined to determine which peer is the next hop toward the originating RP of the SA message. Only if this RPF check succeeds does the SA message get flooded downstream to its peers. This precaution prevents SA messages from looping through the Internet.

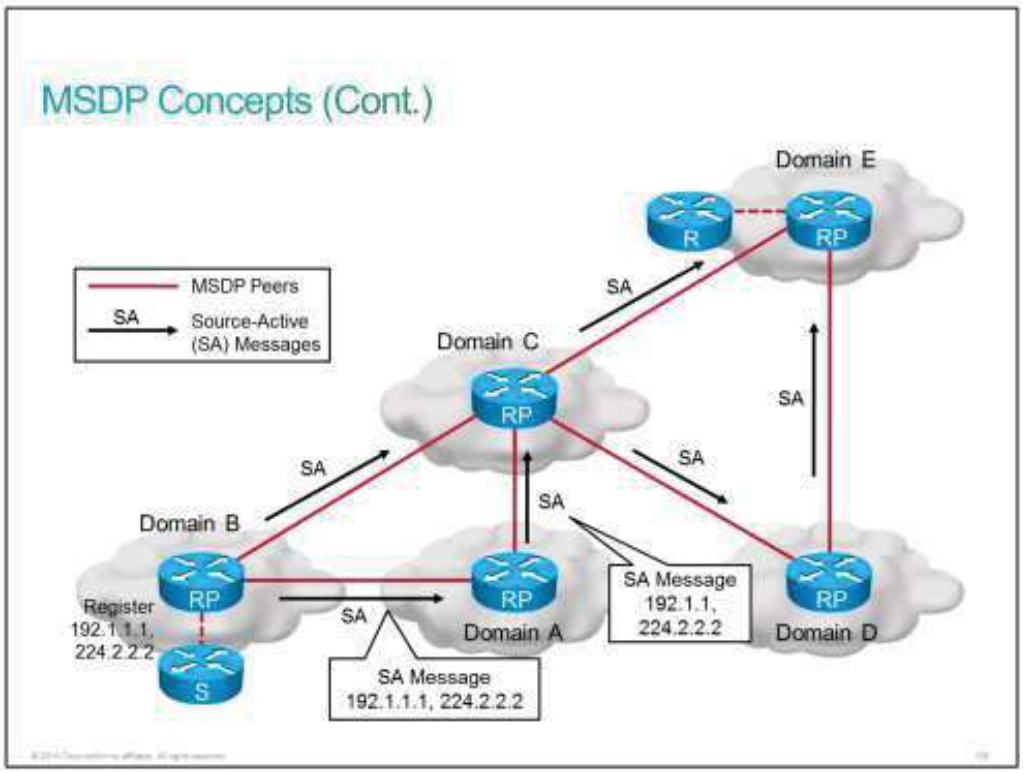
Stub domains (that is, domains with only a single MSDP connection) do not have to perform this RPF check because there is only a single entrance or exit.

MSDP speakers may cache SA messages, but they are typically *not* stored to minimize memory usage. However, by storing SA messages, you can reduce join latency. This reduction is enabled because an RP does not have to wait for the arrival of periodic SA messages when the first receiver joins the group. Instead, the RP can scan its SA cache to immediately determine which sources are active and then send appropriate (S,G) joins.

Noncaching MSDP speakers can query caching MSDP speakers in the same domain for information on active sources for a group.



https://t.me/learningnets




This figure shows an example of PIM-SM domains A through E. Each domain has an RP, which is also an MSDP speaker. The solid lines between RPs represent the MSDP peer sessions via TCP, not actual physical connectivity between the domains. The physical connectivity between the domains is not shown in this figure.

Assume that a receiver in domain E joins multicast group 224.2.2.2. This join causes the DR that is labeled R to send a (*,G) join for this group to the RP.

This activity builds a branch of the shared tree from the RP in domain E to the DR as shown.

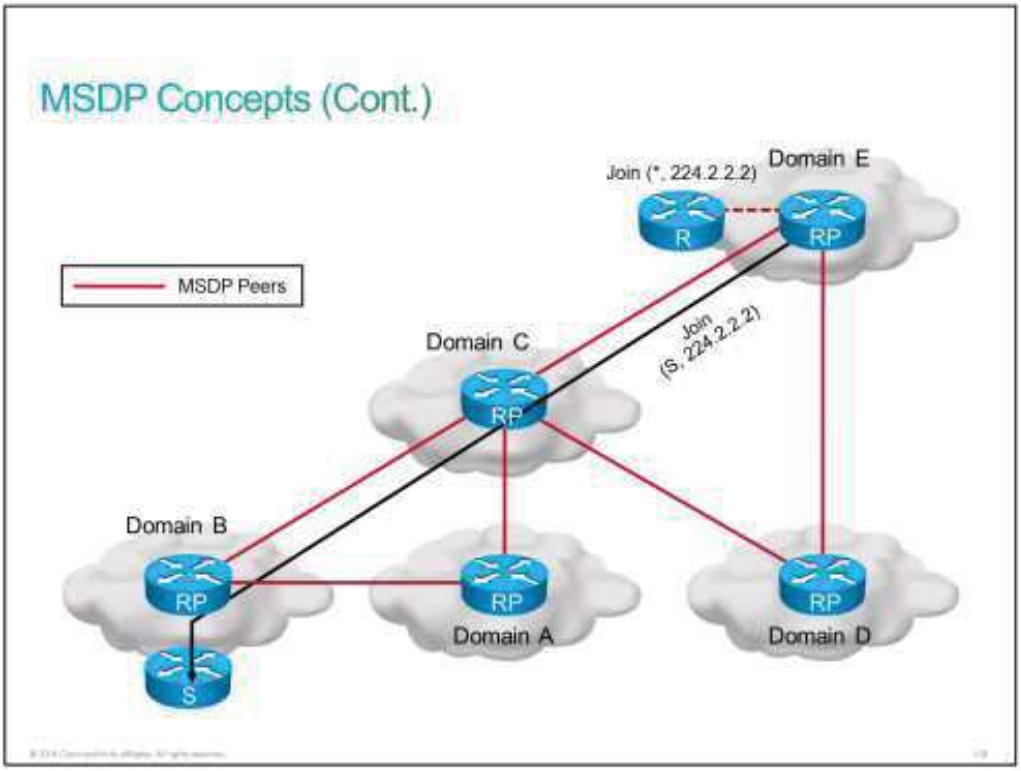
When a source goes active in domain A, the first-hop router S sends a PIM register message to the RP. This informs the RP in domain A that a source is active in the local domain. The RP responds by originating an (S,G) SA message for this source and sending it to its MSDP peers in domains A and C. (The RP will continue to periodically—every 60 seconds—send these SA messages as long as the source remains active.)

When the RPs in domains A and C receive the SA messages, the RPF information is checked. Then, the SA messages are forwarded downstream to the MSDP peers of the domain E and D RPs.

The SA message traveling from domain A to domain C will fail the RPF check at the domain C RP (MSDP speaker). It will then be dropped because domain C has direct connectivity to domain B, where the source exists. However, the SA message arriving at domain C from domain B will pass the RPF check and be processed and forwarded to domains D and E and to A, but the SA from domain C to domain A will fail the RPF check and will be dropped. Similarly, the SA from domain D to domain E will fail the RFP check and will be dropped.

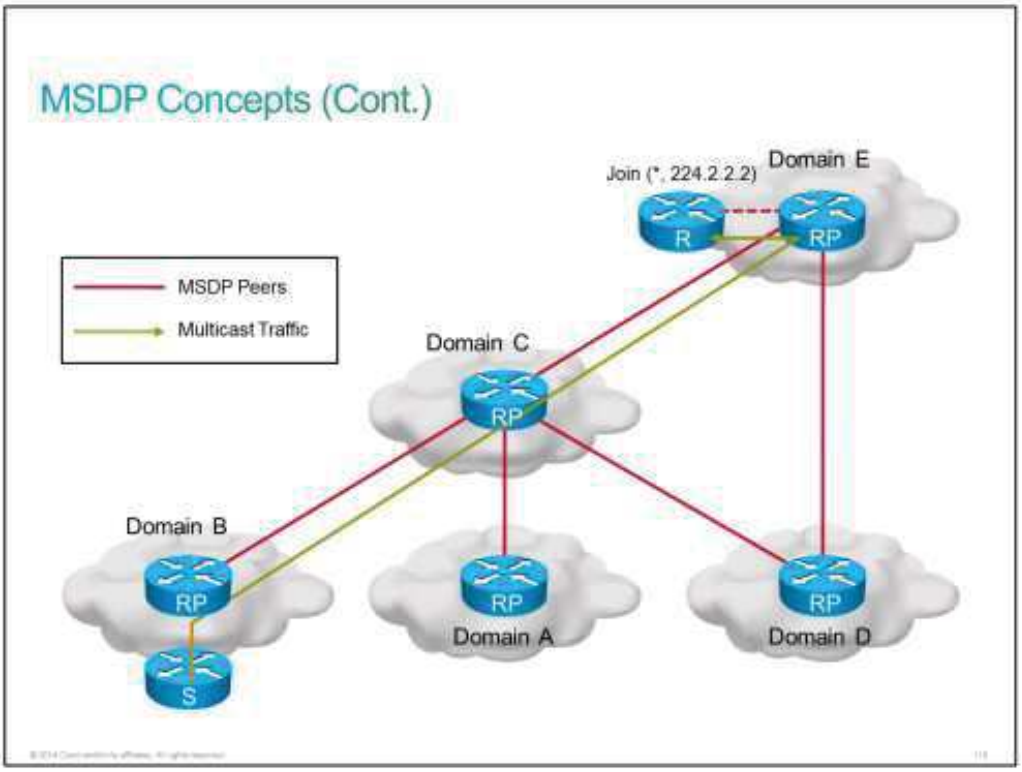


https://t.me/learningnets




Once the SA message arrives at the RP (MSDP speaker) in domain E, the RP sees that it has an active branch of the shared tree for group 224.2.2.2. The RP responds to the SA message by sending an (S,G) join toward the source in domain B.

The (S,G) join will follow the multicast interdomain routing path from the RP in domain E to the source S in domain B. This multicast interdomain routing path is not necessarily the same path that is used by the MSDP connections. The MP-BGP or BGP tables are used to determine the direction in which the (S,G) join is propagated toward the source (more precisely, toward the first-hop router in a domain where the source is active).






https://t.me/learningnets

Once the (S,G) join message reaches the first-hop router in domain B, (S,G) traffic begins to flow to the RP in domain E via the SPT. This activity is shown in the figure.

The (S,G) traffic will follow the path of the source tree that was built by the (S,G) join from the RP in domain E to source S.

Once the (S,G) traffic reaches the last-hop router R in domain E, the last-hop router may optionally send an (S,G) join toward the source in order to bypass the RP in domain E.

The default SPT switchover threshold in PIM-SM for all groups is 0. This means that the switchover will happen immediately after the first multicast packet arrives. Depending on the physical topology of the domain, the shared tree path and the shortest tree path may overlap.

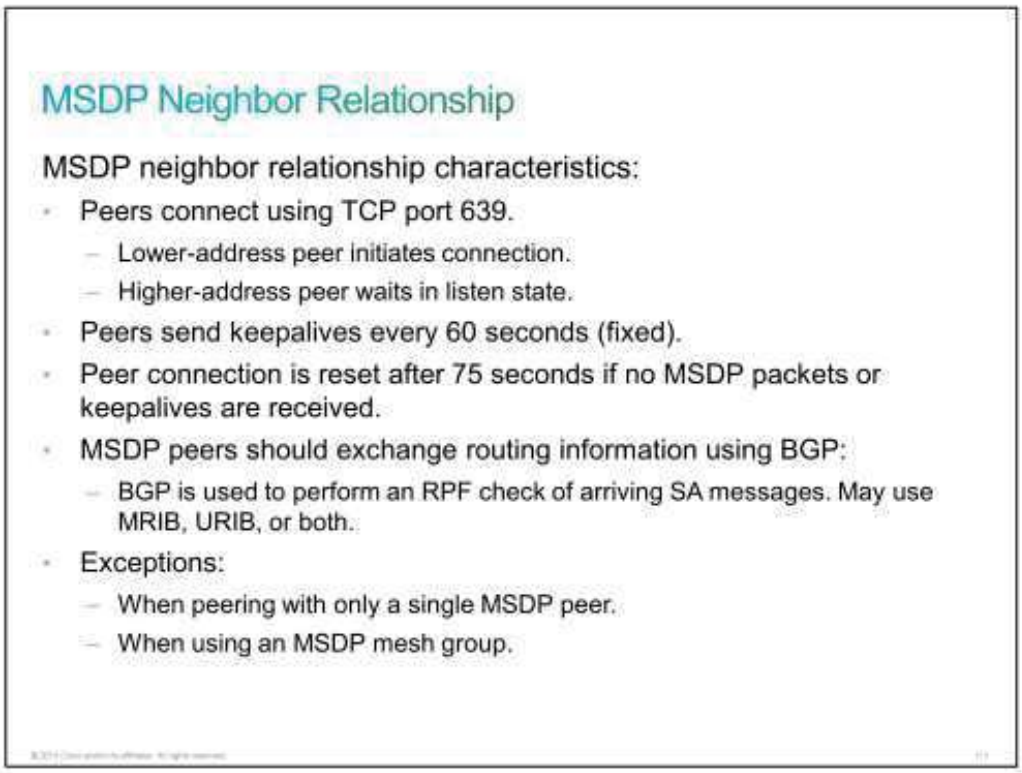
After the SPT switchover occurs in a last-hop router (router R), the (S,G) join from router R reaches source S. The (S,G) traffic from source S will now flow via the SPT to router R where the multicast traffic flows directly from the RP in domain C to router R. The multicast traffic no longer needs to go through the RP in domain E.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# MSDP Neighbor Relationship

Like BGP, MSDP establishes neighbor relationships with other MSDP peers using a TCP session to port 639. MSDP peers send keepalive messages every 60 seconds for a fixed period. The arrival of data performs the same function as the keepalive messages—that is, it keeps the session from timing out. If no keepalive messages or data are received for 75 seconds, the TCP connection is reset and reopened.




MSDP relies upon BGP path information to learn the MSDP topology for the SA peer RPF check. MSDP speakers must run BGP or MP-BGP. This requirement exists because the mechanism that performs the peer RPF check in an SA message uses AS path information that is contained in the MRIB or in the URIB.

There are some special cases where the requirement to perform a peer RPF check on the arriving SA message is suspended. This is the case when there is only a single MSDP peer connection or if the default MSDP peer (default MSDP route) is configured. Also, when MSDP mesh groups are in use, the peer RPF check is skipped for SA messages that arrive from the mesh group members. In these cases, BGP or MP-BGP is not necessary.

| | |
|---|---|
| **Note** | An MSDP mesh group is a group of MSDP speakers that have fully meshed MSDP connectivity between one another. Any SA messages received from a peer in a mesh group are not forwarded to other peers in the same mesh group. The MSDP mesh group has to be configured on each MSDP peer that should be a member of the mesh group. |

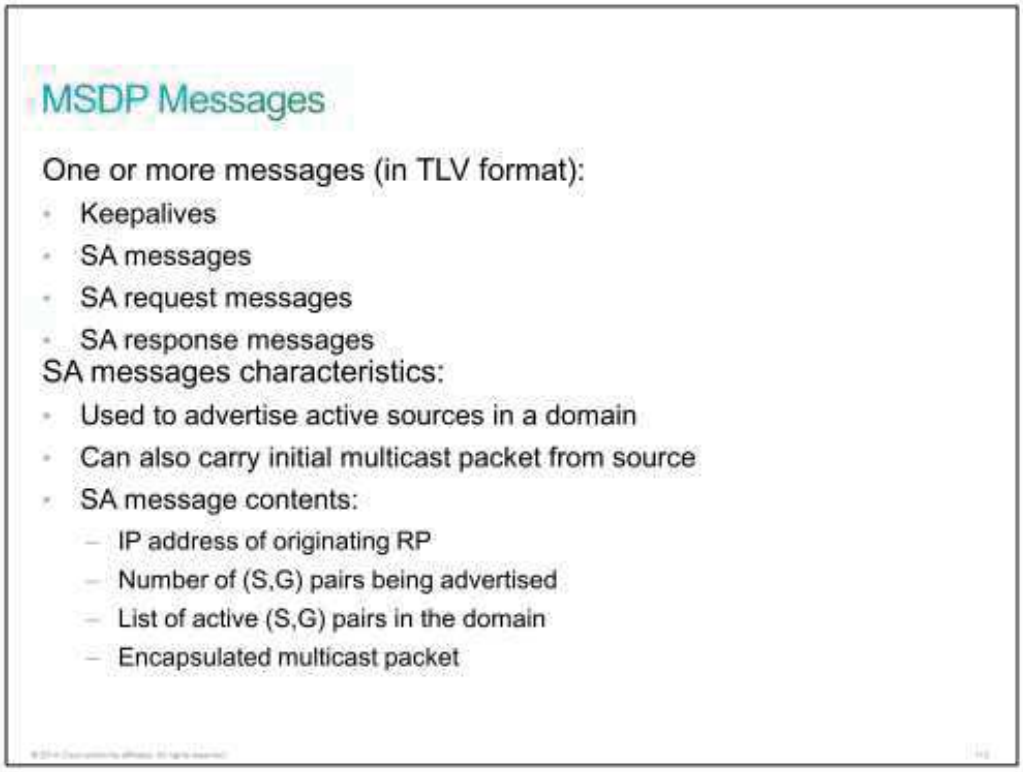


https://t.me/learningnets

# MSDP Messages

SA messages are used to advertise active sources in a domain. In addition, these messages may contain the initial multicast data packet that was sent by the source. Carrying this first data packet in the initial SA message helps to manage the bursty source problem, such as low-rate announcements from a multicast SDR.

SA request messages help decrease join latency, and are used to query the MSDP peers that cache SA messages. Instead of waiting for a periodic SA message, the active sources can be requested as soon as a new group joins a shared tree in a domain.




There are various MSDP messages that are used between MSDP peers. Most of the messages are exchanged between peers only. SA messages are flooded throughout the Internet in a peer-RPF fashion (the RPF check is done to prevent SA looping). There are four basic MSDP message types, each encoded in their own TLV format:

- MSDP keepalive
- SA
- SA request
- SA response

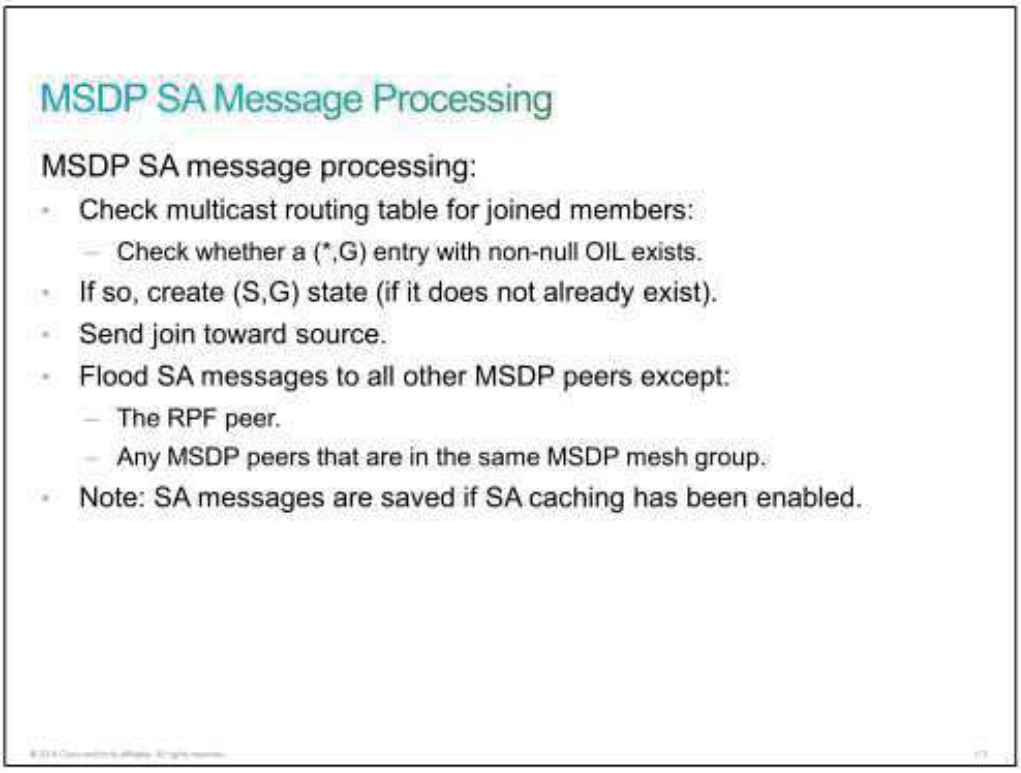
SA response messages are used in response to SA request messages. They are similar in structure to SA messages.



https://t.me/learningnets

# MSDP SA Message Processing

The MSDP SA messages are processed only after successfully passing the peer-RPF check (otherwise they are simply discarded).




A router takes the following steps whenever it processes an MSDP SA message:

- Using group address G of the (S,G) pair in the SA message, the router locates the associated (*,G) entry in the multicast routing table.
- If the (*,G) entry is found and its OIL is not empty, then there are active receivers in the PIM-SM domain for the source that is advertised in the SA message. The following action occurs: the router creates an (S,G) entry for the advertised source and marks it with the flag M (MSDP-created entry).
- If the (S,G) entry does not already exist, the router immediately triggers an (S,G) join toward the source in order to join the source tree.
- The router floods the SA message to all other MSDP peers with the following exceptions:
  - The MSDP peer from which the SA message was received.
  - Any MSDP peers that are in the same MSDP mesh group as the router.

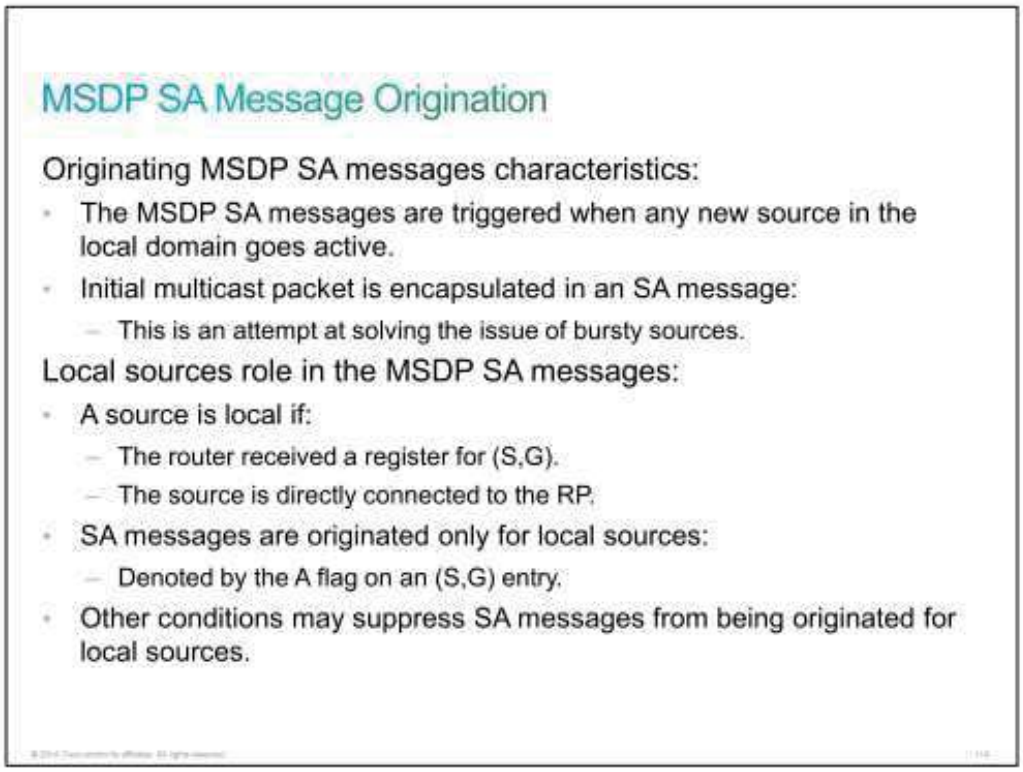
The router does not store SA messages locally unless SA caching has been enabled on the router. In most cases, network administrators enable SA caching in order to improve network debugging capabilities.



https://t.me/learningnets

# MSDP SA Message Origination

The MSDP SA messages are used to advertise active sources in a domain. They are sent periodically every 60 seconds and are normally generated by the RP that is also an MSDP speaker. Additionally, these MSDP SA messages may contain the initial multicast data packet that was sent by the source.




SA messages are triggered by an RP (assuming MSDP is also configured on the RP) when any new source goes active within the local PIM-SM domain.

When a source in the local PIM-SM domain initially goes active, it causes the creation of the (S,G) state in the RP with the A flag set. New sources are detected by the RP by either:

- The receipt of a PIM register message
- The arrival of the first (S,G) packet from a directly connected source

The initial multicast packet sent by the source (either encapsulated in the register message or received from a directly connected source) is encapsulated in the initial SA message in an attempt to solve the problem of bursty sources. SA messages are originated with TTL set to a maximum possible value of 255.

When the multicast packet is encapsulated into an SA message, it could be carried over several hops (in an MSDP session) and thus bypass normal TTL handling. If the TTL thresholds are set in a network to maintain the diameter of multicast traffic, encapsulation in SA messages requires additional filters to be established.

Encapsulating an initial IP multicast packet to an SA message ensures that no multicast packets are lost. This benefit is especially important when managing bursty sources. Because these sources send only one or a few packets, it is possible for sources to be stopped before an SPT is built across domains.

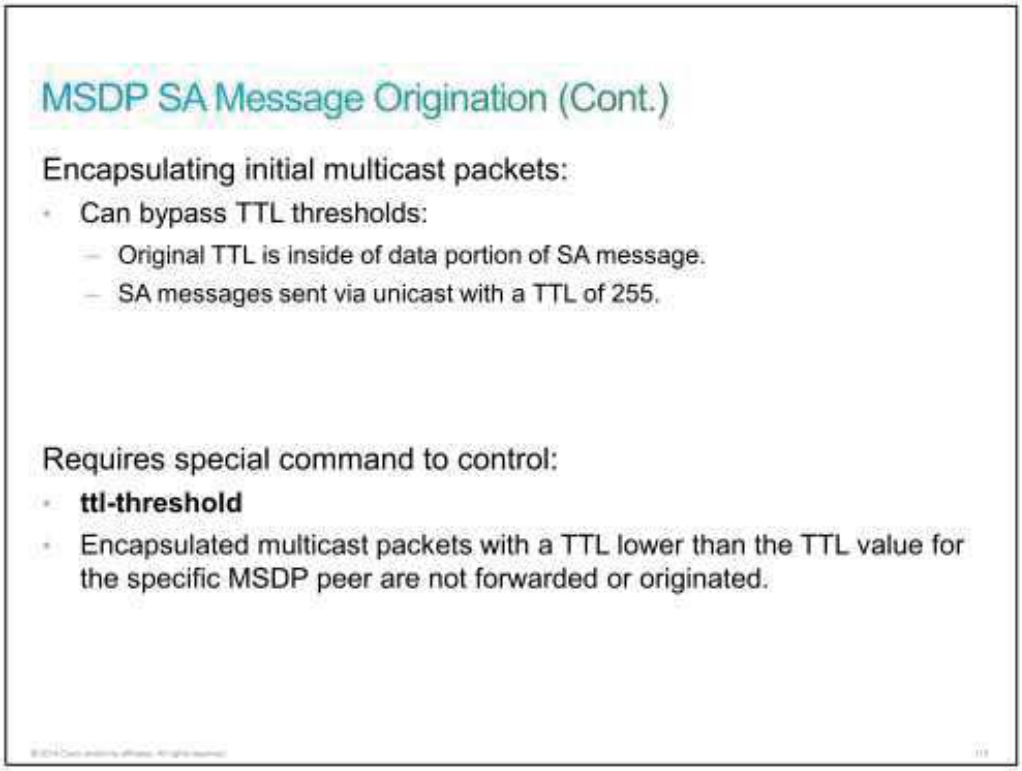
A local source that is advertised by the MSDP router in an SA message is defined as a source when either of the following is true on the MSDP router: it is an RP, or it received a PIM register message.



https://t.me/learningnets

SA messages contain the IP address of the originating RP, the number of (S,G) pairs, and the list of active (S,G) pairs in a domain. A local source is denoted by the A flag being set in the (S,G) entry on the RP or border router. This flag indicates that the source is a candidate for SA advertisement by the RP (MSDP router) to other MSDP peers.

There could also be other conditions that may suppress SA messages from being originated for local sources. More details will be covered in the material that follows.




A TTL threshold problem can be introduced by the encapsulation of the initial multicast packet of a source in an SA message. Because the multicast packet is encapsulated inside of the unicast SA message (whose TTL equals 255), its TTL is not decremented as the SA message travels to the MSDP peer. Furthermore, the total number of hops that the SA message traverses can be drastically different than the number traversed by a normal multicast packet. This situation exists because multicast and unicast traffic may follow completely different paths to the MSDP peer and, therefore, the remote PIM-SM domain. This scenario can result in TTL thresholds being violated by this encapsulated packet.

The solution to this problem is to configure a TTL threshold that is associated with any multicast packet that is encapsulated in an SA message sent to a particular MSDP peer. This configuration can be accomplished by using the Cisco IOS/IOS XE command **ip msdp ttl-threshold** *peer-address ttl* or the Cisco IOS XR **ttl-threshold** *ttl* command. The following example shows Cisco IOS XR syntax:

```
router msdp
 peer peer-address
  ttl-threshold ttl
```

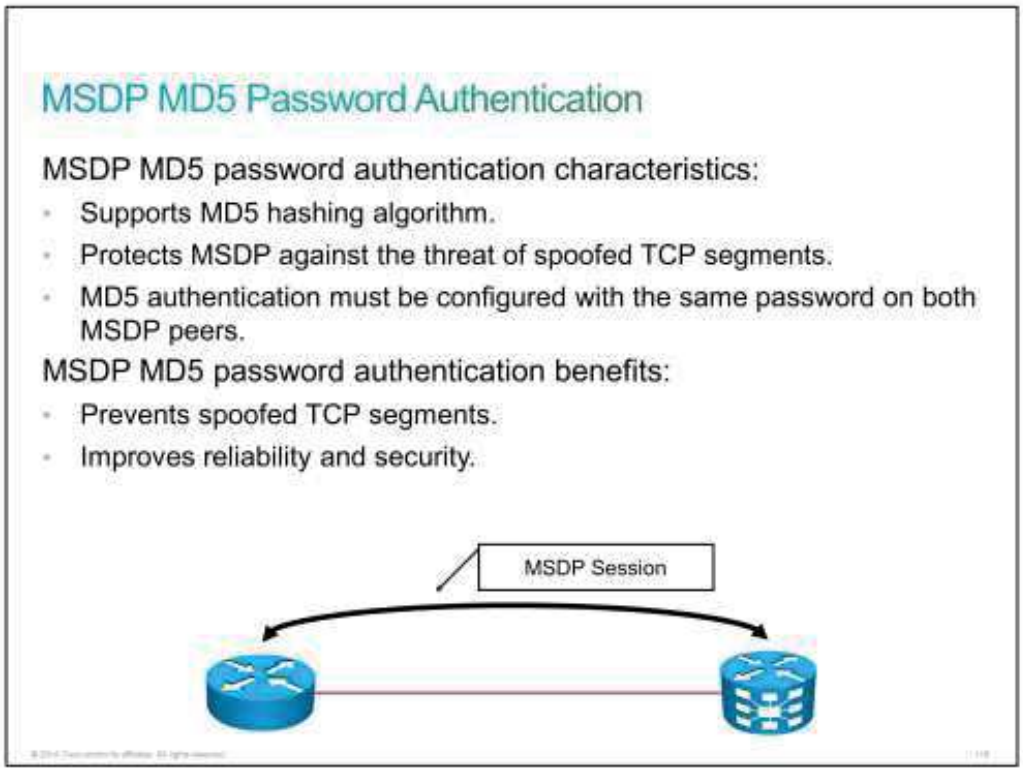
This command prevents any multicast packet whose TTL is below the TTL value (*ttl*) from being encapsulated in an SA message sent to the MSDP peer whose IP address is *peer-address*.



https://t.me/learningnets

# MSDP MD5 Password Authentication

The MSDP MD5 password authentication feature is an enhancement to support MD5 signature protection on a TCP connection between two MSDP peers. This feature provides added security by protecting MSDP against the threat of spoofed TCP segments being introduced into the TCP connection stream.




Developed in accordance with RFC 2385, the MSDP MD5 password authentication feature is used to verify each segment sent on the TCP connection between MSDP peers.

When MD5 authentication is enabled between two MSDP peers, each segment sent on the TCP connection between the peers is verified. MD5 authentication must be configured with the same password on both MSDP peers, or the connection between the peers will not be made.

These are the benefits of MSDP MD5 password authentication:

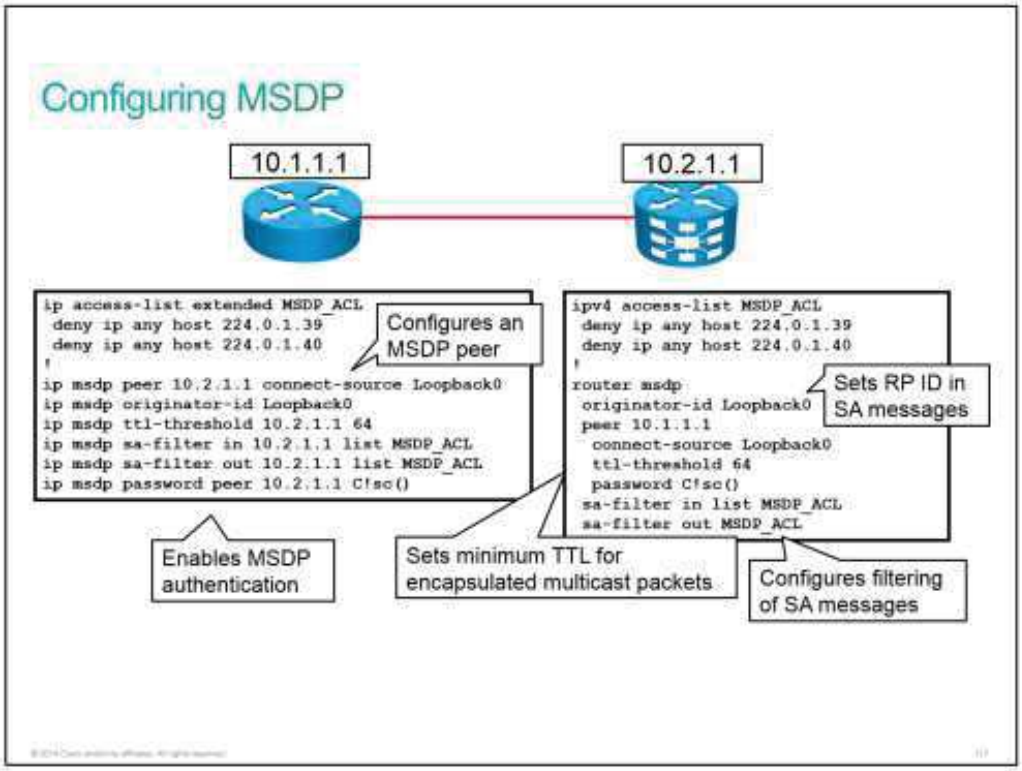
- It protects MSDP against the threat of spoofed TCP segments being introduced into the TCP connection stream.
- It uses the industry-standard MD5 algorithm for improved reliability and security.



https://t.me/learningnets

# Configuring MSDP

To configure an MSDP peer, use the Cisco IOS/IOS XE **ip msdp peer** global configuration command or the Cisco IOS XR **peer** router MSDP command. You can also use the optional **connect-source** command to specify the IP address from which the MSDP TCP session is established.




An MSDP speaker that originates an SA message can use the IP address of an interface as the RP address in the SA message. To enable this, use the Cisco IOS/IOS XE **ip msdp originator-id** global configuration command or the Cisco IOS XR **originator-id** router MSDP command.

To configure an incoming filter list for SA messages received from the specified MSDP peer, use the Cisco IOS/IOS XE **ip msdp sa-filter in** global configuration command or the Cisco IOS XR **sa-filter in** router MSDP command. If this command is not configured, no incoming messages are filtered, and all SA messages are accepted from the peer. If the command is configured but no access list, route map, or RPL is specified, all source and group pairs from the peer are filtered.

To configure an outgoing filter list for SA messages that are sent to the specified MSDP peer, use the Cisco IOS/IOS XE **ip msdp sa-filter out** global configuration command or the Cisco IOS XR **sa-filter out** router MSDP command. If this command is not configured, no outgoing messages are filtered, and all SA messages that are received are forwarded to the peer. If the command is configured but no access list, route map, or RPL is specified, all source and group pairs are filtered. Filtering SA messages can break the flood-and-join mechanism.

To limit which multicast data packets are sent to an MSDP peer in SA messages, use the Cisco IOS/IOS XE **ip msdp ttl-threshold** global configuration command or the Cisco IOS XR **ttl-threshold** router MSDP command.

Use the Cisco IOS/IOS XE **ip msdp password peer** command or the Cisco IOS XR **password** router MSDP command to enable MD5 authentication for TCP connections between two MSDP peers.



https://t.me/learningnets

# Verifying MSDP

To verify MSDP peers, use the **show msdp summary** command. This command will display MSDP peers and the state of the MSDP connection.

## Verifying MSDP

```
RP/0/RSP0/CPU0:PE5# show msdp summary
Out of Resource Handling Enabled
Maximum External SA's Global : 20000
Current External Active SAs : 0
MSDP Peer Status Summary
Peer Address AS  State   Uptime/     Reset  Peer   Active  Cfg.Max     TLV
                         Downtime    Count  Name   SA Cnt  Ext.SAs     recv/sent
10.6.1.1     0   Up      00:23:32    1      7      0       0           30/48
```

- Displays summary of MSDP peers.

## Verifying MSDP (Cont.)

```
RP/0/RSP0/CPU0:PE5# show msdp sa-cache
MSDP Flags:
E - set MRIB E flag , L - domain local source is active,
EA - externally active source, PI - PIM is interested in the group,
DE - SAs have been denied. Timers age/expiration, Cache Entry:
(192.168.156.60, 226.1.1.1), RP 10.6.1.1, MBGP/AS 0, 00:00:04/00:02:25
  Learned from peer 10.6.1.1, RPF peer 10.6.1.1
  SAs recvd 1, Encapsulated data received: 0
    grp flags: none, src flags: EA
(10.6.1.1, 226.1.1.1), RP 10.6.1.1, MBGP/AS 0, 00:00:04/00:02:25
  Learned from peer 10.6.1.1, RPF peer 10.6.1.1
  SAs recvd 1, Encapsulated data received: 0
    grp flags: none, src flags: EA
```

- Displays states learned from MSDP peers.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

To verify active multicast sources that were learned from MSDP peers, use the **show msdp sa-cache** command.



https://t.me/learningnets

# Summary

This topic summarizes the key points that were discussed in this lesson.

## Summary

- Interdomain multicast traffic has different requirements as intradomain forwarding.
- GLOP is based on temporary allocation of 233.0.0.0/8.
- SSM in interdomain scenario eliminates the need for RPs and shared trees.
- MSDP announces active sources to the neighboring domains.
- MSDP uses interdomain source trees.

## Summary (Cont.)

- MSDP uses TCP to exchange information.
- SA messages are used to advertise active sources in a domain.
- They are triggered when any new source in the local domain goes active.
- MSDP messages can be authenticated using MD5 authentication.
- MSDP is configured separately from other multicast configuration.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# Lesson 4: Identifying Rendezvous Point Distribution Solutions

## Overview

This lesson offers an overview of RP distribution solutions. It explains the drawbacks of manual RP configuration, and it describes the Auto-RP and the BSR mechanisms. The lesson also introduces the concept of Anycast RP, which works in combination with the MSDP.

Upon completing this lesson, you will be able to:

- Identify disadvantages of using static RP configuration
- Identify two dynamic RP discovery mechanisms (Auto-RP and BSR)
- Discuss the placement of the RP in the network
- Describe the Auto-RP mechanism
- Describe Auto-RP candidate RPs
- Describe Auto-RP mapping agents
- Describe other (not candidate RP and not mapping agent) routers
- Discuss Auto-RP configuration
- Discuss Auto-RP troubleshooting steps
- Discuss Auto-RP scoping
- Discuss how to secure Auto-RP using a boundary
- Describe the BSR mechanism
- Describe BSR candidate RPs



https://t.me/learningnets

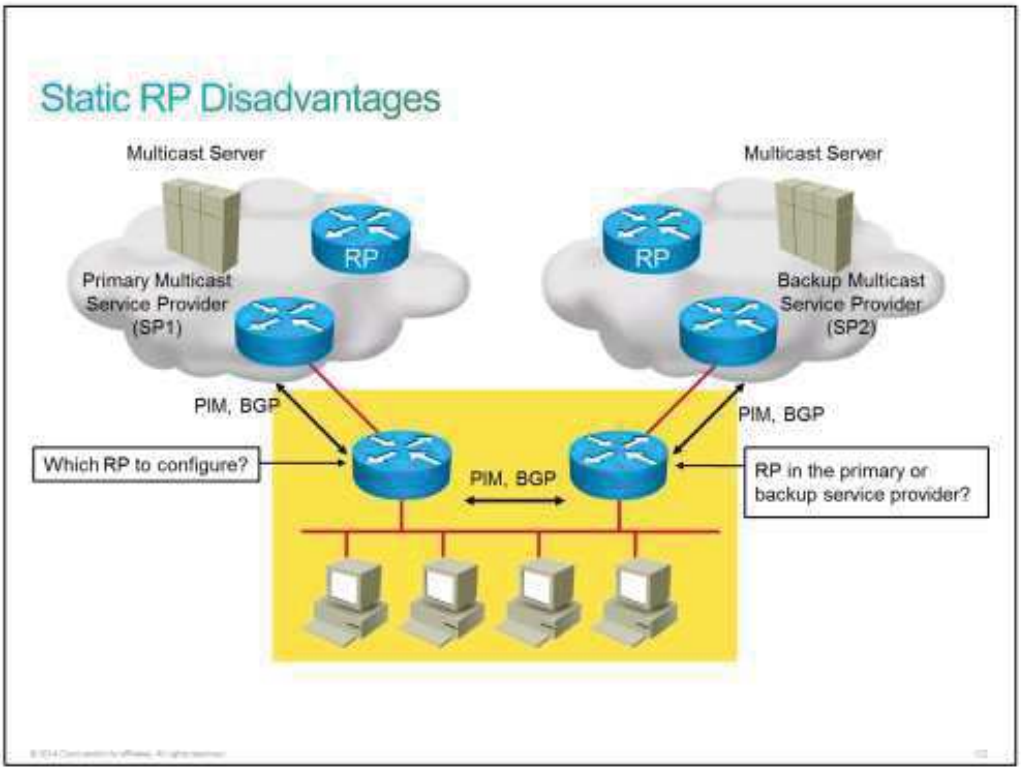
- Describe BSRs
- Describe the BSR election process
- Describe PIMv2 routers that are not BSRs or candidate RPs
- Describe the BSR advertisement process
- Discuss BSR configuration
- Discuss BSR verification
- Discuss BSR troubleshooting steps
- Discuss BSR hop-by-hop flooding
- Discuss how to constrain BSR messages
- Describe using Anycast RP in conjunction with MSDP
- Describe an example of using Anycast RP in conjunction with MSDP
- Discuss Anycast RP configuration
- Discuss Anycast RP configuration guidelines

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# Static RP Disadvantages

Addresses that are within the 224.0.0.0 to 224.0.0.255 range are considered link-local multicast addresses. Packets that are multicast to these addresses are always transmitted with a TTL of 1, so that they do not leave the local link.




Manual RP information configuration does not scale to large environments. There are a several problems that are associated with its static nature:

- The initial configuration of many routers is cumbersome.
- Any change to the static configuration has a negative impact on the maintenance effort, which in the long term turns out to be even more troublesome than the initial provisioning.
- Static RP information configuration does not support many scenarios with redundant RPs, in which the network reachability changes, as in the redundant setup in the figure.

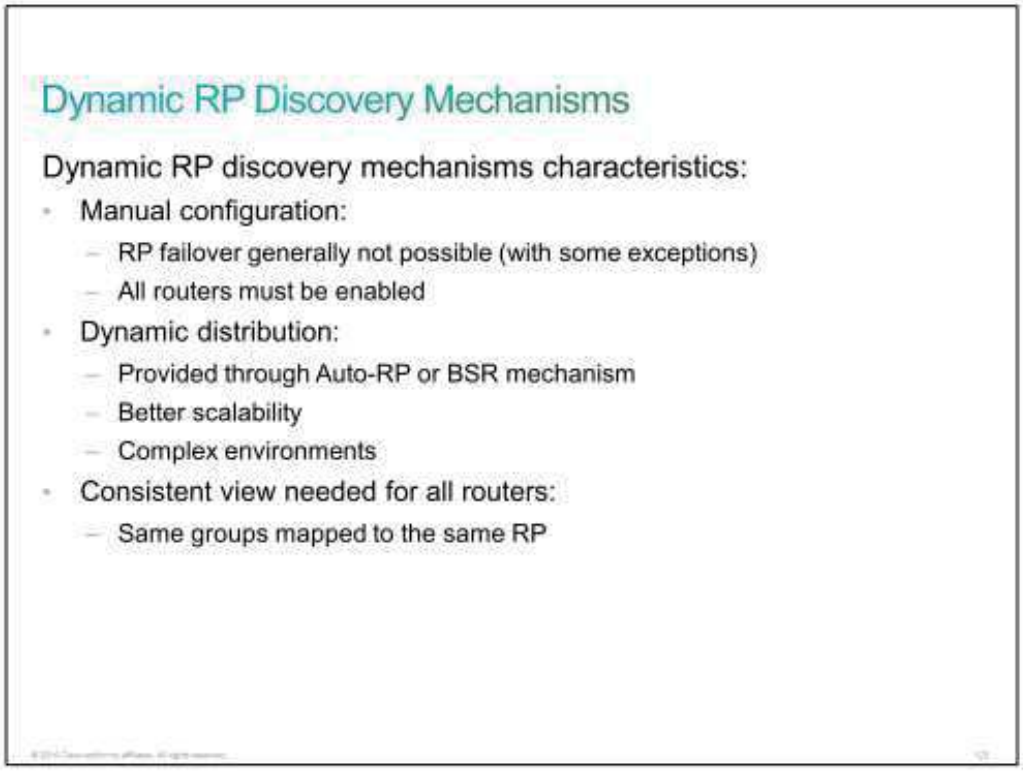
In the figure, the enterprise has two uplinks to two service providers that offer multicasting services. The providers multicast to the same groups, but have independent sources and RPs. The enterprise has PIM and BGP. These protocols are peering with both service providers to exchange routing information and build multicast distribution trees. SP1 is the primary provider, while SP2 is the backup, from which the multicast traffic should be received only if the connection to SP1 fails. Static RP information configuration does not support such requirements, because the RP of SP1 has a different address than the RP of SP2. If the RP of SP1 were configured on the enterprise routers, the RP selection would never fall back to the RP of SP2, if the primary link went down.



https://t.me/learningnets

# Dynamic RP Discovery Mechanisms

Dynamic RP discovery mechanisms: Auto-RP and BSR are described below:




In addition to the manual RP information configuration method, two mechanisms of dynamic RP discovery have been developed to improve scalability in large environments:

- Auto-RP, which is a method designed by Cisco.
- BSR, which is a standards-based mechanism.

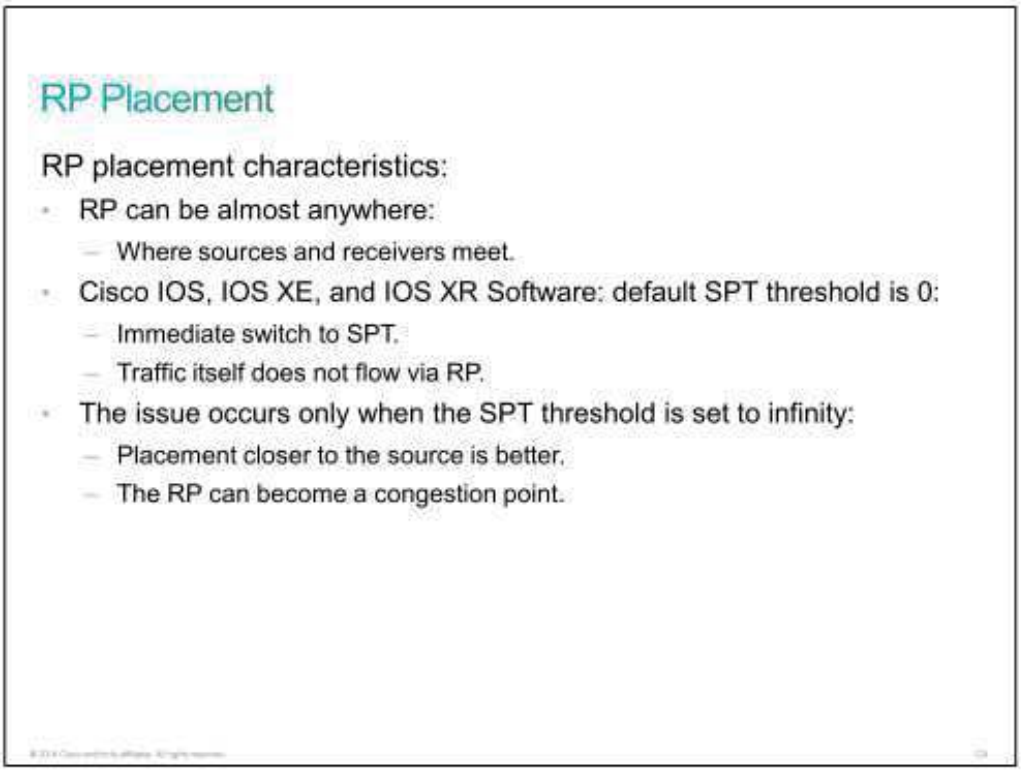
No matter which mechanism is used, the designer must ensure that all routers have a consistent view. In other words, the same range of multicast groups has to be mapped to the same RP.



https://t.me/learningnets

# RP Placement

The decision as to which RP information distribution to use is not directly linked to the question of which routers should be declared as RPs or candidate RPs.




RPs can be almost anywhere. Theoretically, they do not even need to be on the path from the sources to the receiver. On the Cisco IOS, IOS XE, and IOS XR Software platforms, the default SPT threshold is set to 0. This default means that the distribution trees immediately switch to SPT, and the traffic itself does not flow via the RP.

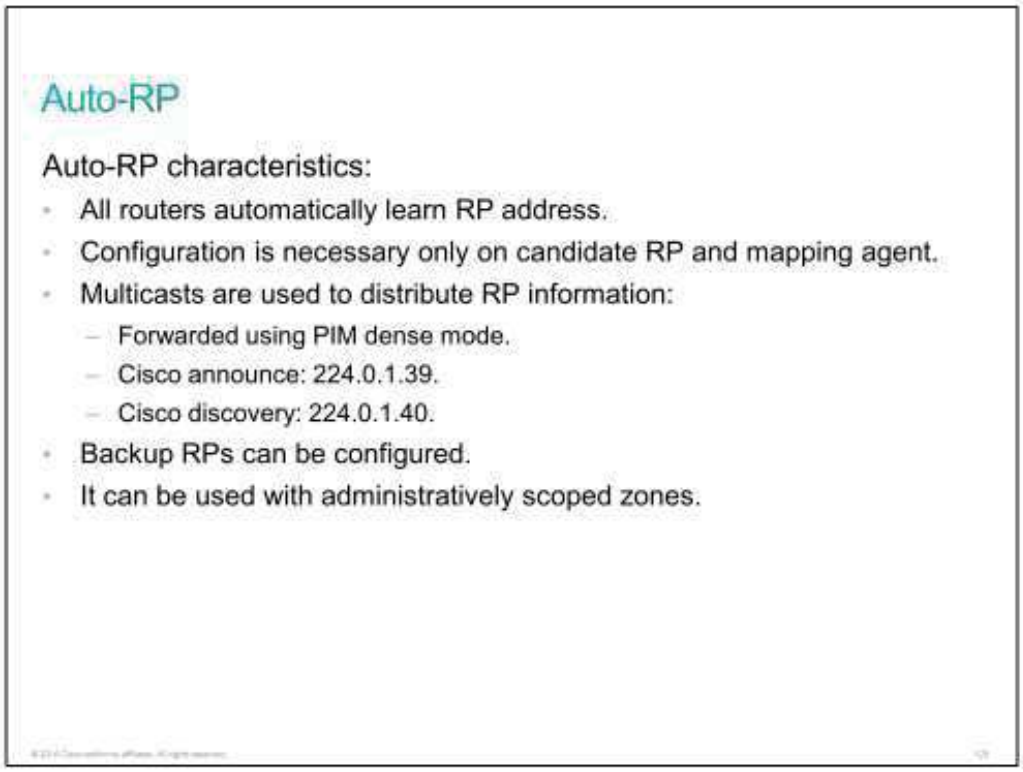
For other values of the SPT threshold, especially infinity, a placement closer to the source is recommended, as the RP can become a congestion point.



https://t.me/learningnets

# Auto-RP

Auto-RP allows all routers in the network to automatically learn group-to-RP mappings, and is enabled in the access, IP edge, and core of the Cisco IP NGN infrastructure layer.




There are no special configuration steps that must be taken, except on the routers that are to function as:

- Candidate RPs
- Mapping agents

Multicast is used to distribute group-to-RP mapping information via two special multicast groups that are assigned with IANA addresses:

- Cisco announce group: 224.0.1.39
- Cisco discovery group: 224.0.1.40

Because multicast is used to distribute this information, a which-occurred-first situation can happen if the 224.0.1.39 and 224.0.1.40 groups operate in sparse mode. Routers would have to know the RP address before they could learn the address of the RPs via Auto-RP messages. Therefore, it is recommended that these groups run in PIM-DM so that this information is flooded throughout the network.

Multiple candidate RPs may be defined so that in the case of an RP failure, the other candidate RP can assume the responsibility of the RP.

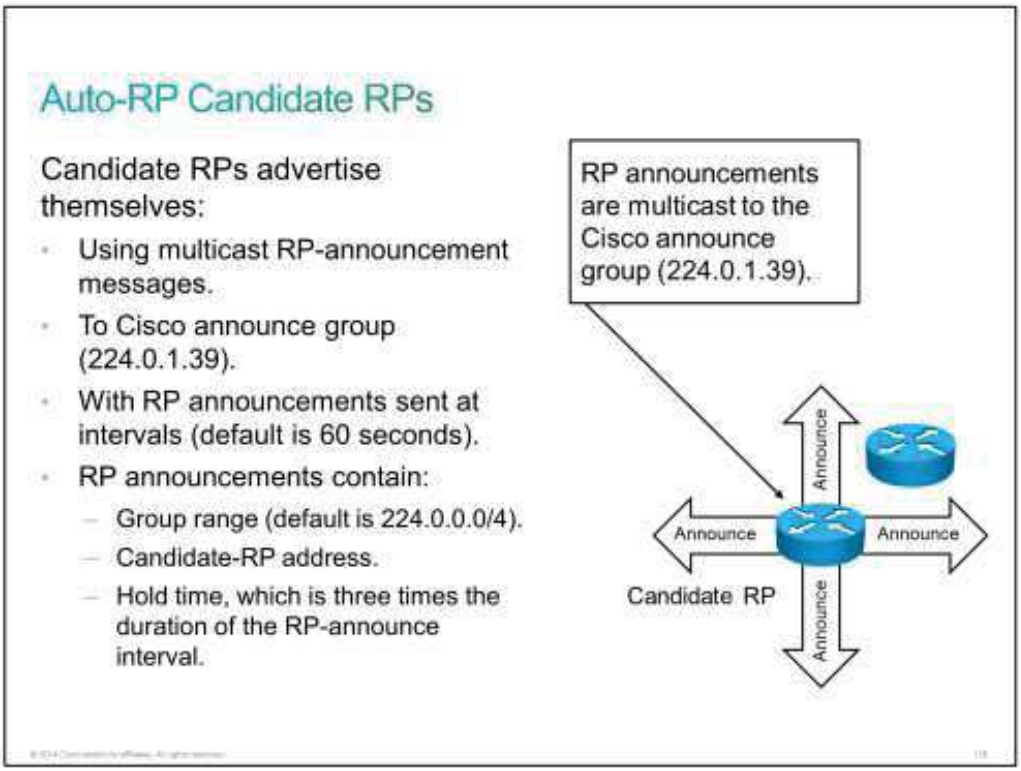
Auto-RP can be configured to support administratively scoped zones, unlike the BSR mechanism, which is explained later. Administratively scoped zones can be important when trying to prevent high-rate group traffic from leaving a campus and consuming too much bandwidth on the WAN links.



https://t.me/learningnets

# Auto-RP Candidate RPs

An Auto-RP candidate RP is sending RP-announcement messages to the Cisco announce group (224.0.1.39). These messages announce the router as being a candidate RP. By default, the messages are sent every 60 seconds.




RP-announcement messages contain:

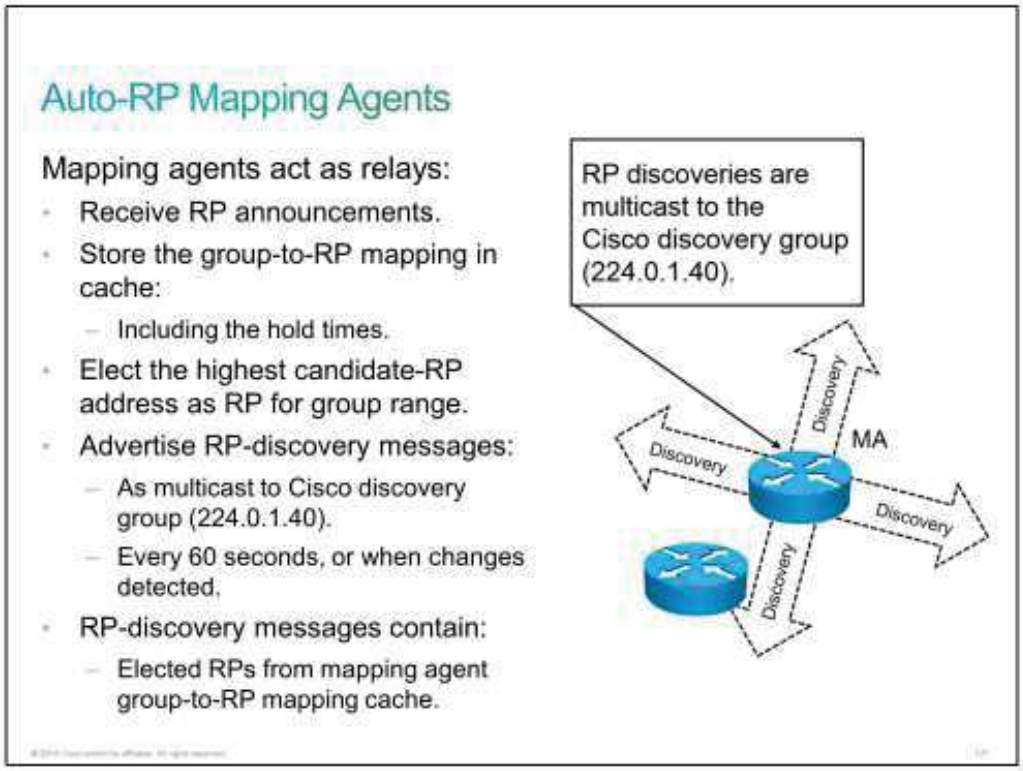
- The group address range (the default is the all-multicast-groups address, or 224.0.0.0/4).
- The IP address of the candidate RP.
- A hold time, which is used to detect when the candidate RP has failed. This hold time is three times the announcement interval (the default is 180 seconds).



https://t.me/learningnets

# Auto-RP Mapping Agents

An Auto-RP mapping agent joins the RP-announcement group (224.0.1.39) in order to receive RP announcements from a candidate RP.




When a mapping agent receives an announcement:

- It saves the announcement in the group-to-RP mapping cache.
- It selects the candidate RP with the highest IP address as the RP for the group range.

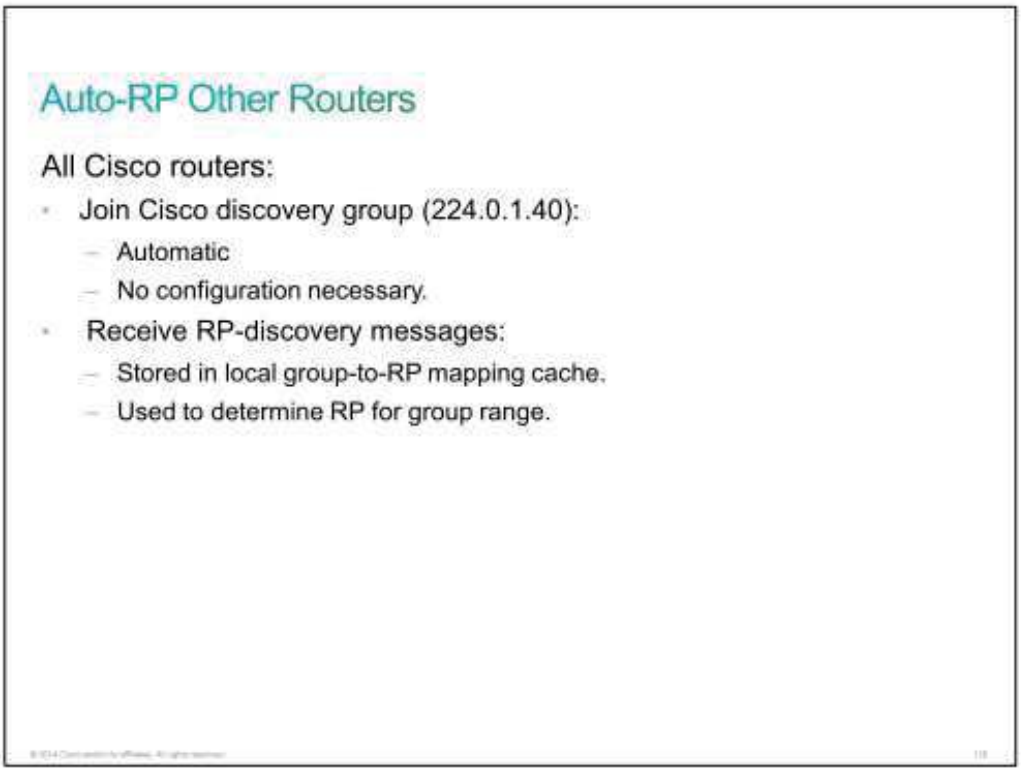
The hold times are used to expire an entry in the cache if a candidate RP fails and is no longer sending periodic candidate-RP announcements. Via RP-discovery messages, the mapping agent periodically sends information on the elected RPs from its group-to-RP mapping cache to all routers in the network. RP-discovery messages are multicast to the Auto-RP-discovery group (224.0.1.40). They are sent every 60 seconds or when a change to the information in the group-to-mapping cache takes place.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# Auto-RP Other Routers

Cisco routers automatically join the Cisco discovery group (224.0.1.40) in order to receive the group-to-RP mapping information being multicast by the mapping agent in the network. No configuration is required to join this group.



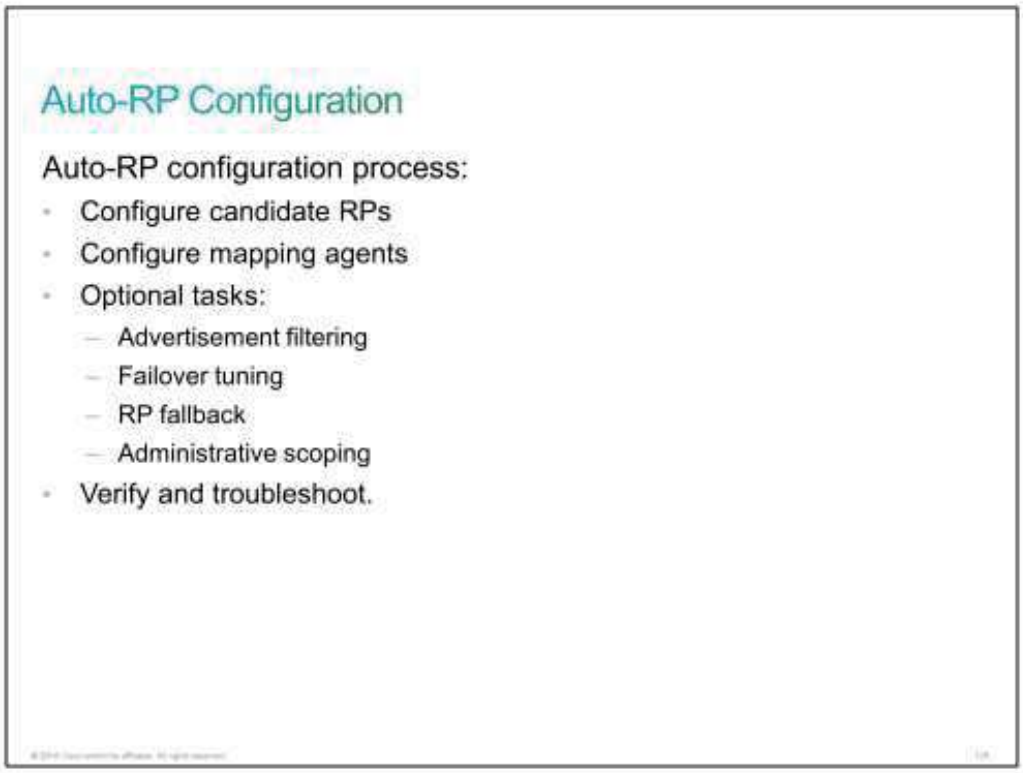

Group-to-RP mapping information that is contained in the RP-discovery messages is stored in the local group-to-RP mapping cache of the router. The router uses this information to map a group address to the IP address of the active RP for the group.



https://t.me/learningnets

# Auto-RP Configuration

To implement Auto-RP, you must configure Auto-RP candidate RPs, which send RP-announcement messages to the Cisco announce group (224.0.1.39). Sent every 60 seconds by default, these messages announce that a router is a candidate for selection as an RP. RP-announcement messages contain the group range (the default is the all-multicast-group range, or 224.0.0.0/4), the IP address of the candidate, and a hold time, which is used to detect when the candidate RP has failed. The hold time duration is three times the announcement interval, which is 180 seconds.




To implement the Auto-RP mechanism, you must configure the following:

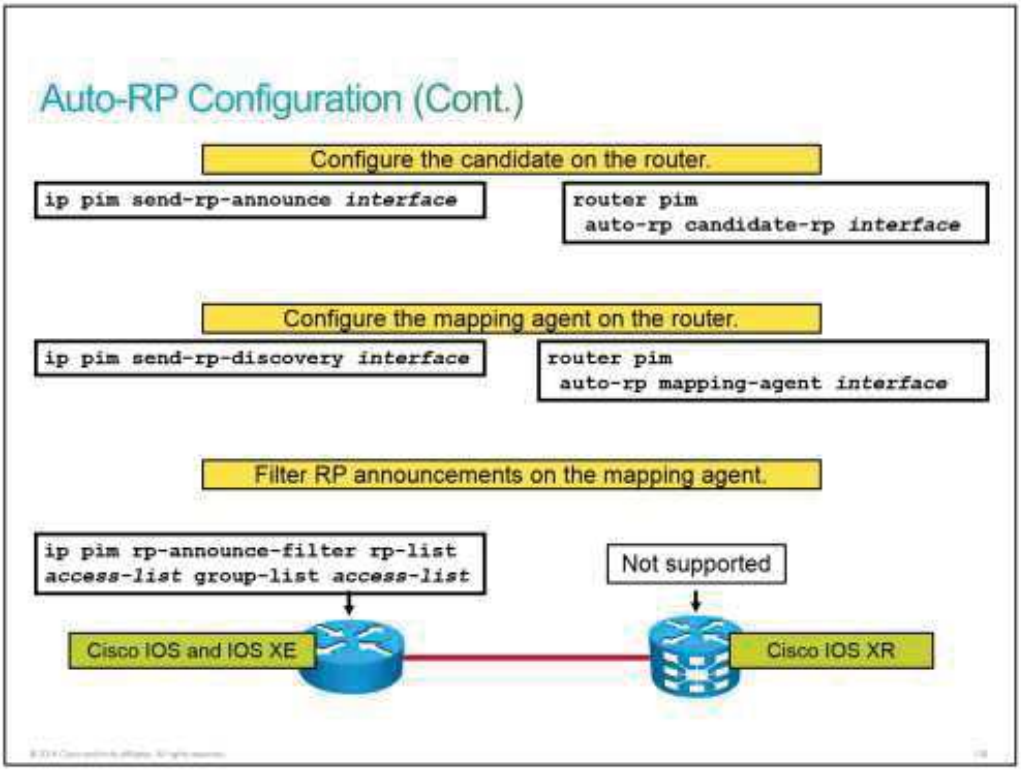
- One or more candidate RPs
- One or more mapping agents

Additionally, these are the Auto-RP tuning options:

- Advertisement filtering
- Failover tuning
- RP fallback
- Scope constraining

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets




For Auto-RP, you must configure one or more Auto-RP mapping agents. Mapping agents join the RP-announcement group (224.0.1.39) in order to receive RP announcements that are sent by all candidate RPs. When mapping agents receive an announcement, they save the announcement in the group-to-RP mapping cache and select the candidate RP with the highest IP address as the RP for the group range. The hold time is used to time out an entry in the cache if a candidate RP fails and is no longer sending periodic candidate-RP announcements. Via RP-discovery messages, mapping agents periodically send information to the elected RPs from their group-to-RP mapping cache to all routers in the network. RP-discovery messages are multicast to the Auto-RP discovery group (224.0.1.40). They are sent every 60 seconds or when a change to the information in the group-to-mapping cache takes place.

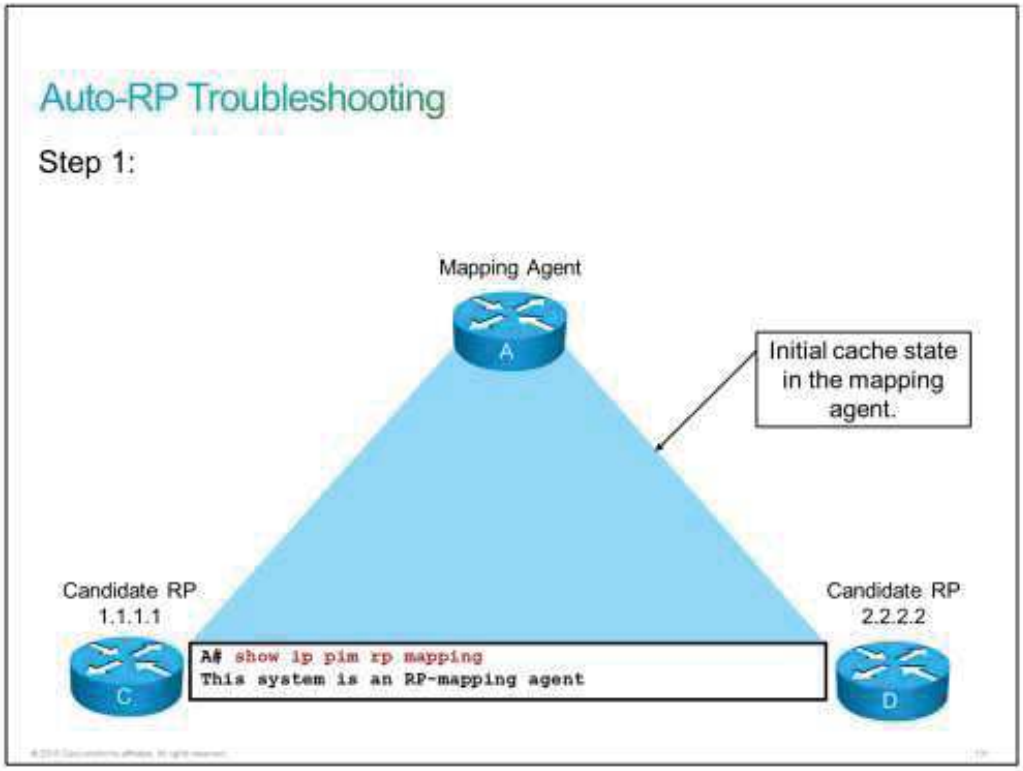
Filtering RP announcements is supported on the Auto-RP mapping agents. Network administrators may wish to configure mapping agents so that they will only accept candidate-RP announcements from well-known routers in the network. This prevents the acceptance of candidate-RP announcements from bogus routers, which may then potentially be selected as the RP.



https://t.me/learningnets

# Auto-RP Troubleshooting

Because other routers in the network will learn the group-to-RP mapping information from the mapping agents, it is important that this information is correct on the mapping agents. If the information is not correct, you must verify that the candidate RPs are configured correctly. You must also verify that the mapping agent properly received the candidate-RP announcements.




If multiple mapping agents are in use, make sure that their group-to-RP mapping information is identical. If it is not, the routers in the network will oscillate between the different RPs selected by the agents. Again, make sure that all the mapping agents are properly receiving Auto-RP announcements from all candidate RPs in the network.

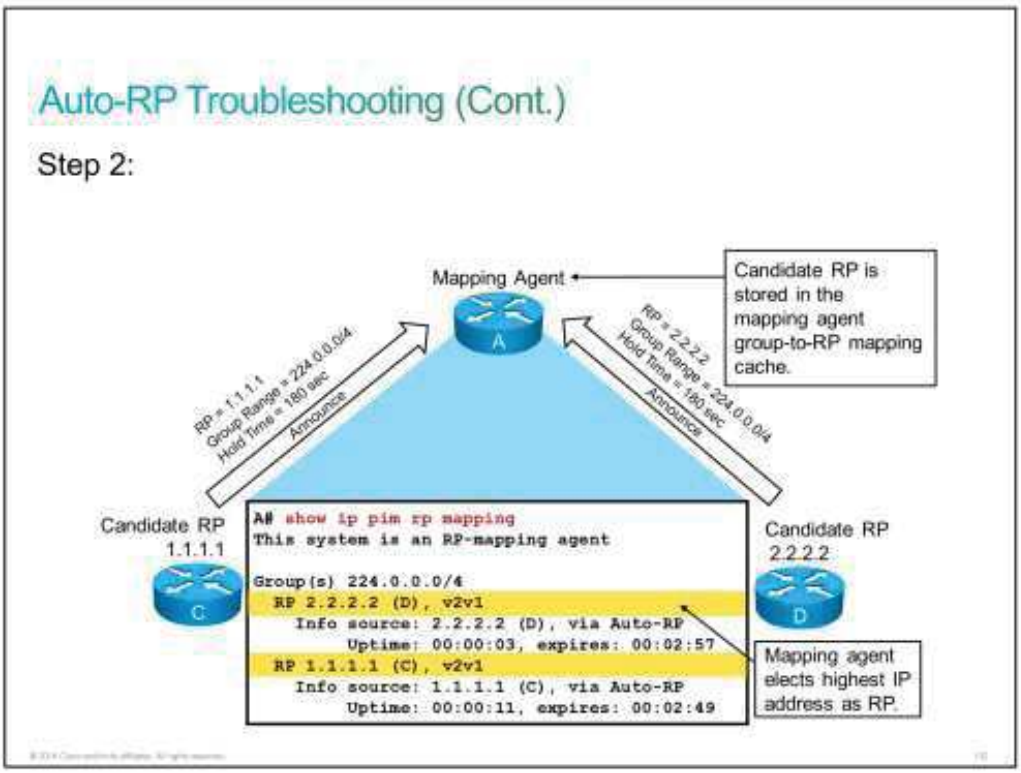
Group-to-RP mapping information should match the information of the mapping agents. If not, you must verify that the router is properly receiving Auto-RP discovery messages from the agents.

Take care when configuring the TTL scope on the RPs and mapping agents, especially with a redundant setup with multiple routers in those roles. If the TTL is too low, the messages may not reach all the destination routers. If the TTL is too high, an overlap will ensue, and you will need to ensure that the information offered by the multiple mapping agents is identical. Another challenge is to constrain the message overflow into other administrative domains, which will be discussed later in this topic.

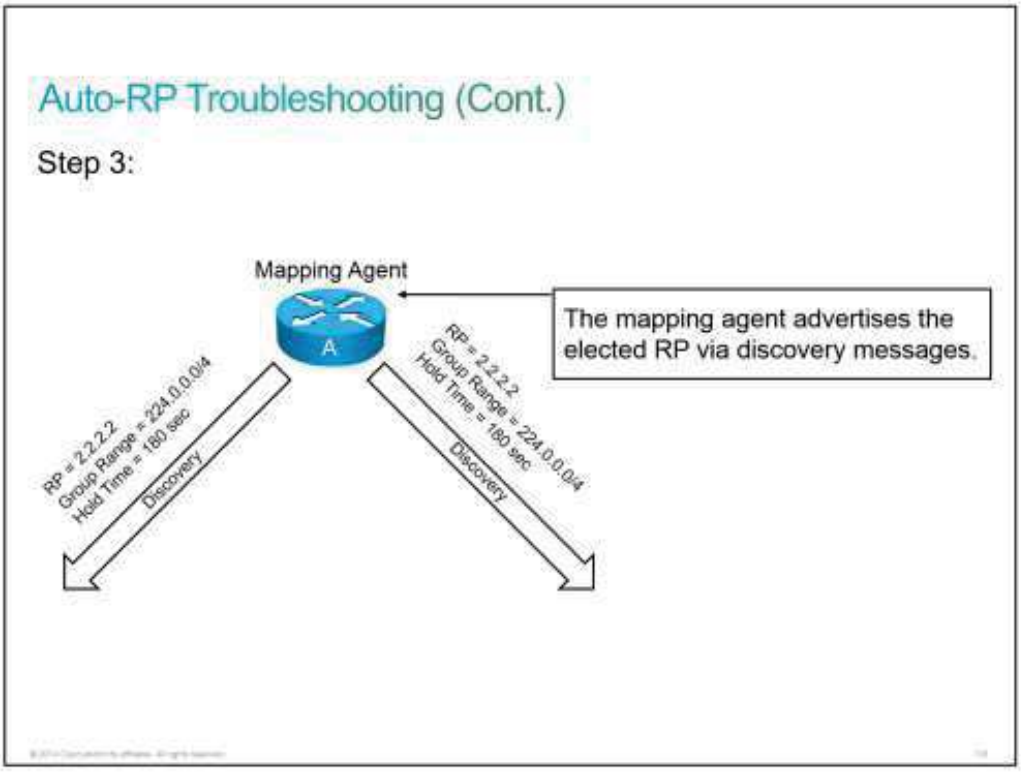
The example in this page and the following pages examine the process in more detail at each step. As shown in this figure, at time zero, the group-to-RP mapping caches in the mapping agents are empty, since no RP announcements have been received. The output of the Cisco IOS/IOS XE **show ip pim rp mapping** command or the Cisco IOS XR **show pim group-map** command shows that router A is a mapping agent and that the group-to-RP mapping cache is empty.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets



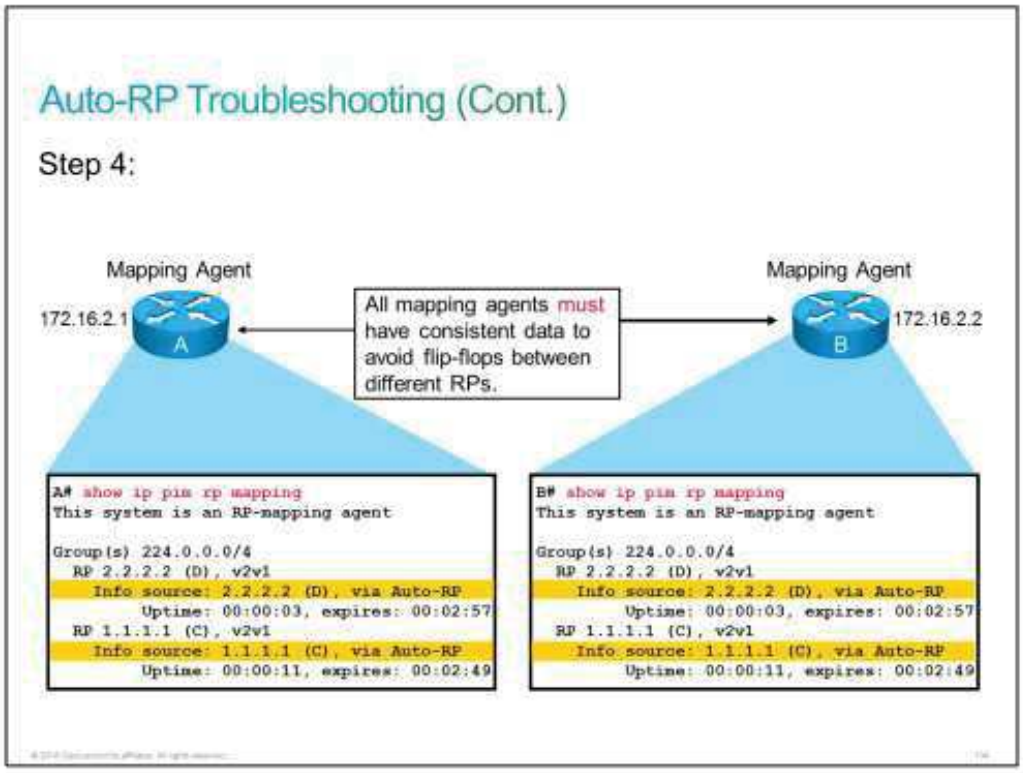

Routers C and D begin sending their RP-announcement messages to advertise themselves as RP candidates for all multicast groups. The mapping agent (router A) receives these RP announcements and stores this information in its group-to-RP mapping cache. The output on the mapping agent now shows both routers C and D as candidates for group range 224.0.0.0/4 (that is, all multicast groups except for the Auto-RP groups, 224.0.1.39 and 224.0.1.40). The mapping agent then elects the candidate RP with the highest IP address as the active RP for the group range.



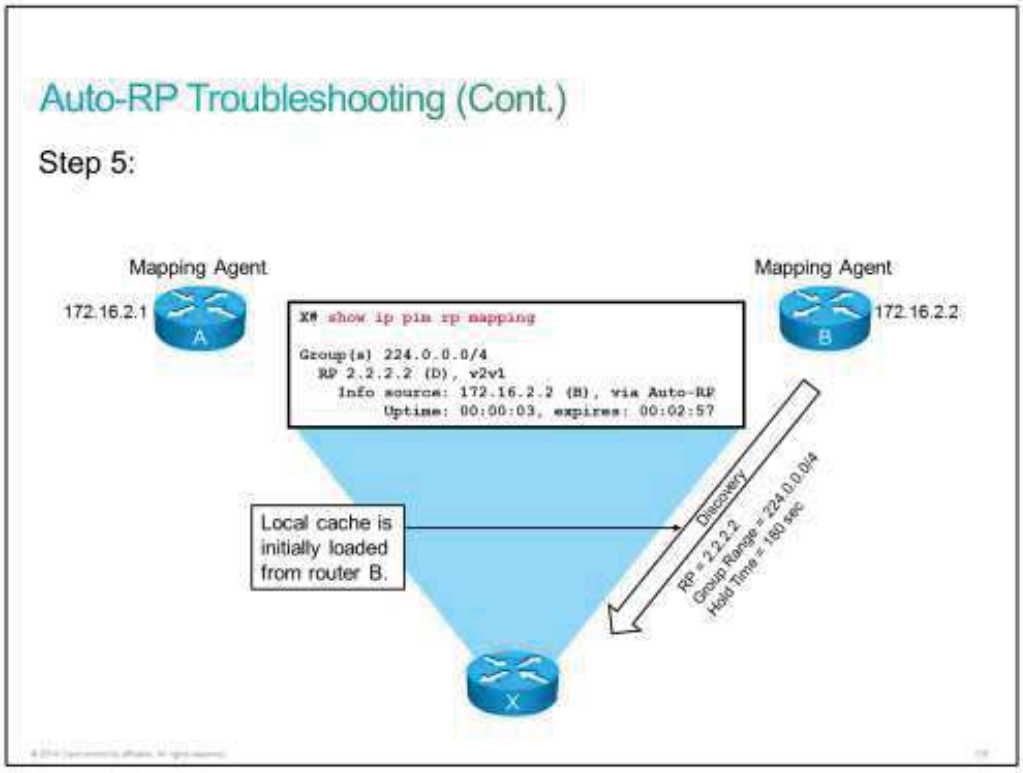



https://t.me/learningnets

The mapping agent begins to advertise the results of the RP election to the rest of the network via Auto-RP discovery messages.




It is critical that all mapping agents in the PIM-SM domain have identical information in their group-to-RP mapping caches. If the information in the mapping caches is not identical, it can cause the routers in the network to flip-flop between two different RPs.

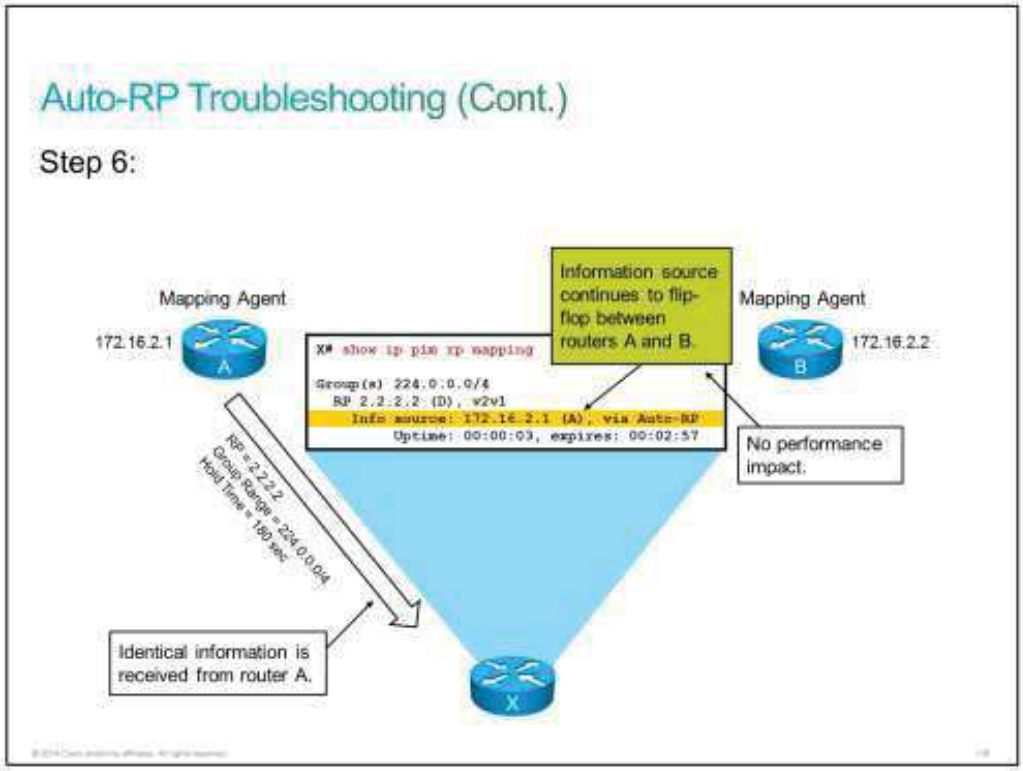



 © 2014 Cisco Systems, Inc.

https://t.me/learningnets

Assume that router B is the first mapping agent to send its RP-discovery message containing the contents of its group-to-RP mapping cache. The routers in the network all receive this RP-discovery message and install the information in their local group-to-RP mapping cache.

The output shows that the 2.2.2.2 router (router D, not shown in this diagram) is currently selected as the RP for group range 224.0.0.0/4 (that is, all multicast groups except the Auto-RP groups). It also shows that this information was most recently received from router B.




Router A sends an RP-discovery message containing the contents of its group-to-RP mapping cache. The routers in the network receive this RP-discovery message and update the information in their local group-to-RP mapping cache. Since both mapping agents are sending identical information, the only thing that will change in the local group-to-RP mapping cache is the source of the information.

The output shows that the 2.2.2.2 router (router D, not shown in this diagram) is still selected as the RP for group range 224.0.0.0/4 (that is, all multicast groups except the Auto-RP groups). However, the data reflects that this information was most recently received from router A. The flip-flop of the information source in the group-to-RP mapping cache of the local router has negligible impact on the router.

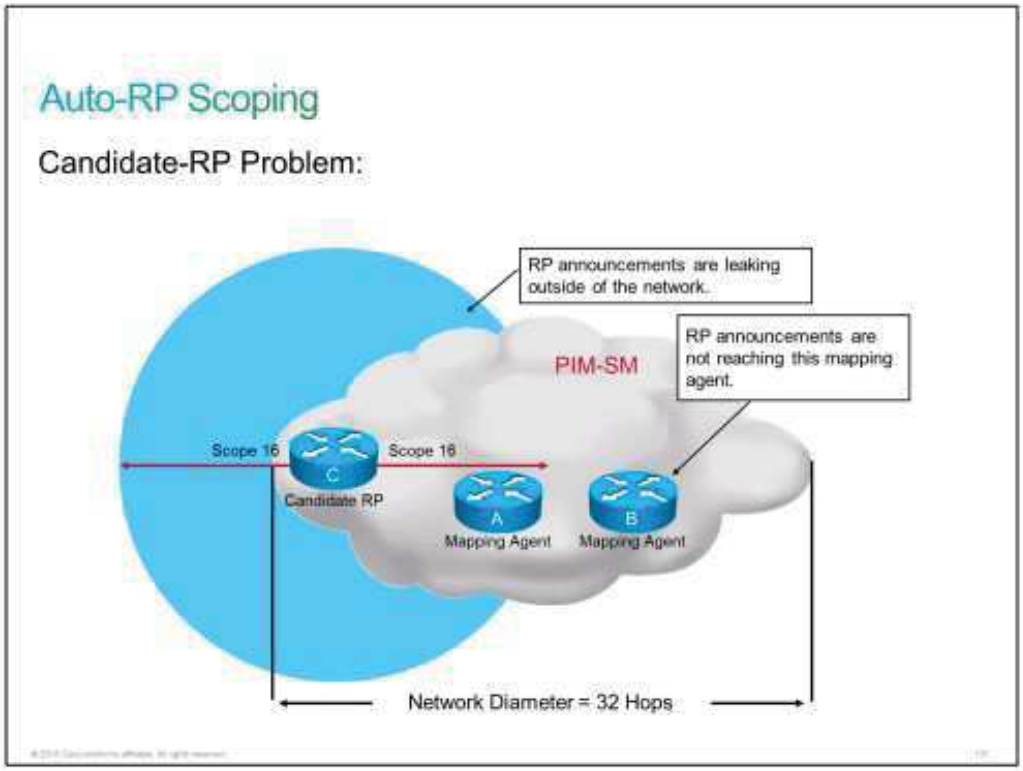


https://t.me/learningnets

# Auto-RP Scoping

Care must be taken in the selection of the TTL scope of RP-announcement messages that are sent by candidate RPs to ensure that the RP announcement messages reach all the required mapping agents.

Care must also be taken in the selection of the TTL scope of RP-discovery messages that are sent by mapping agents to ensure that the messages reach all PIM-SM routers in the network.

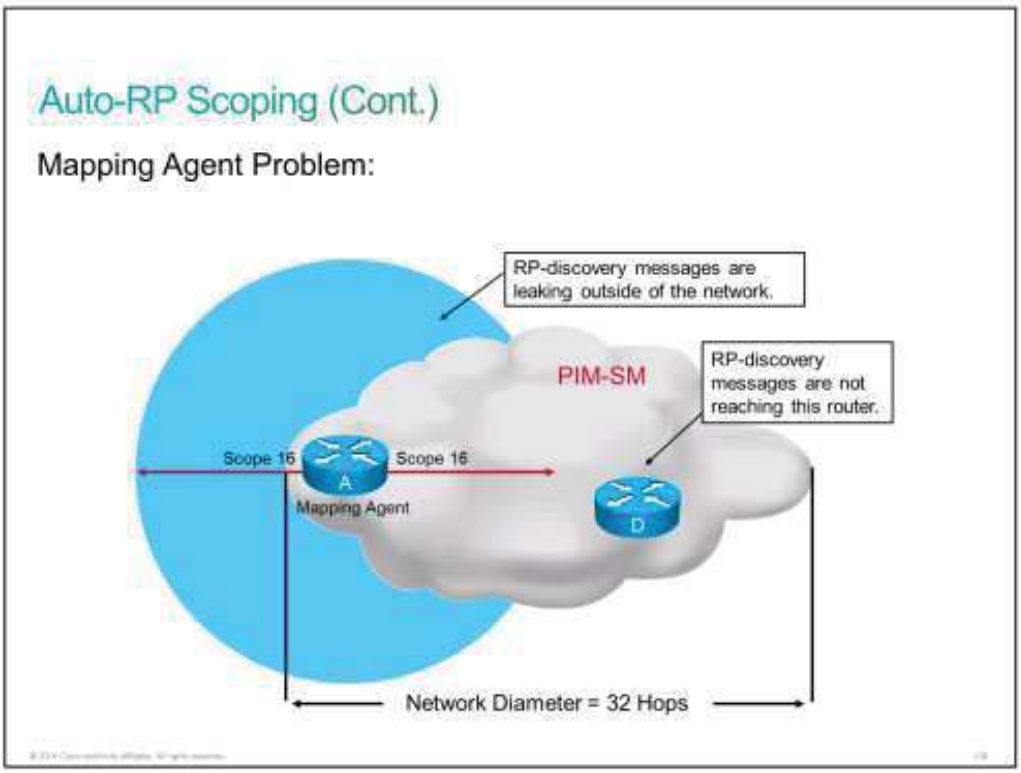



In the figure, an arbitrary scope of 16 was used in the Cisco IOS and IOS XE **ip pim send-rp-announce** command or the Cisco IOS XR **auto-rp candidate-rp** router PIM command on the candidate-RP router. However, the maximum diameter of the network is greater than 16 hops, and in this case one mapping agent (router B) is farther away than 16 hops. As a result, this mapping agent does not receive the RP-announcement messages from the candidate RP. This lapse can cause the two mapping agents to have different information in their group-to-RP mapping caches. If this event occurs, each mapping agent will advertise a different router as the RP for a group, which will have disastrous results.

Also, the candidate RP is fewer than 16 hops way from the edge of the network. This arrangement can cause RP-announcement messages to leak into adjacent networks and produce Auto-RP problems in those networks.



https://t.me/learningnets




In the figure, an arbitrary scope of 16 was used in the Cisco IOS and IOS XE **ip pim send-rp-discovery** command or the Cisco IOS XR **auto-rp mapping-agent** router PIM command on the mapping agent. However, the maximum diameter of the network is greater than 16 hops, and in this case, at least one router (router D) is farther away than 16 hops. As a result, this router does not receive the RP-discovery messages from the mapping agent. This lapse can result in the router having no group-to-RP mapping information. If this event occurs, the router will attempt to operate in dense mode for all multicast groups while other routers in the network work in sparse mode.

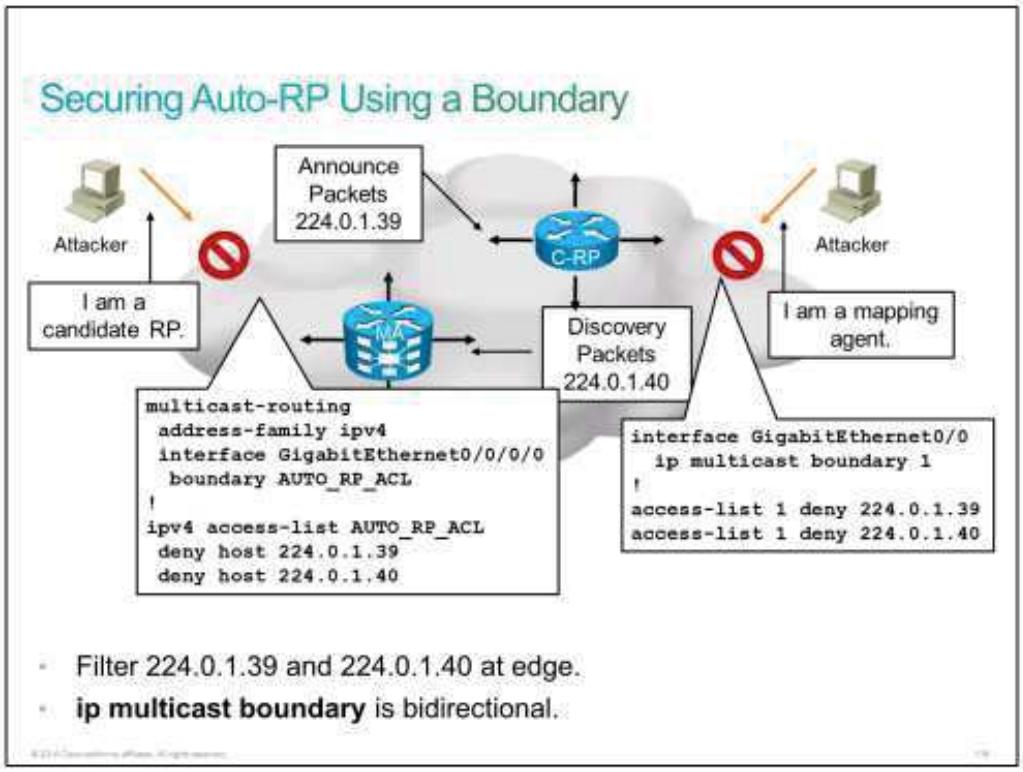
Also, the mapping agent is fewer than 16 hops away from the edge of the network. This arrangement can cause RP-discovery messages to leak into adjacent networks and produce Auto-RP problems in those networks.



https://t.me/learningnets

# Securing Auto-RP Using a Boundary

The forwarding of Auto-RP packets is always enabled, and currently not configurable. This default can present a particular security exposure in the case of Auto-RP.




Both 224.0.1.39 and 224.0.1.40 are forwarded in PIM-DM to avoid the multicast routers having to know the RP for a group when that group is used to distribute RP information. This usage is the only recommended use of PIM-DM.

You can set up an administratively scoped boundary on an interface for Auto-RP multicast group addresses in order to prevent multicast traffic for Auto-RP groups being sent to or received from another multicast domain. To set up the boundary, use the Cisco IOS and IOS XE **ip multicast boundary** command or the Cisco IOS XR **boundary** command. A standard access list defines the range of addresses affected. When a boundary is set up, no multicast data packets are allowed to flow across the boundary from either direction. The boundary allows the same multicast group address to be reused in different administrative domains.

| | |
|---|---|
| **Note** | The Cisco IOS and IOS XE `ip multicast boundary` command or the Cisco IOS XR `boundary` command can be used in general to administratively scope multicast traffic within one domain. |

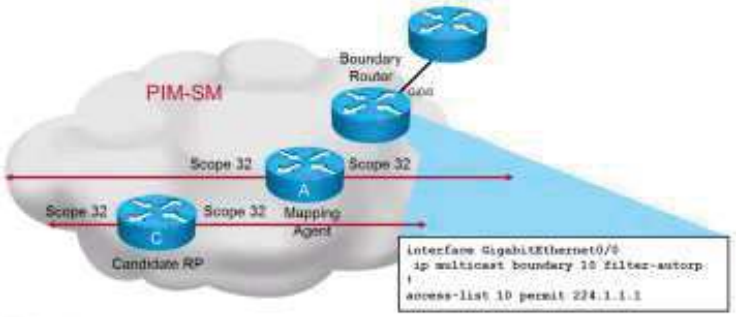


https://t.me/learningnets

## Securing Auto-RP Using a Boundary (Cont.)

Securing Auto-RP using a boundary characteristics:

- The **filter-autorp** option filters Auto-RP discovery and announcement messages.
- It permits an Auto-RP group range announcement only if all addresses in the Auto-RP group range are permitted by access list.
- This is important, because RP designation can be configured only on RPs, not on the leaf routers.




You can configure the Cisco IOS and IOS XE **ip multicast boundary filter-autorp** keyword to examine and filter Auto-RP discovery and announcement messages at the administratively scoped boundary. Any Auto-RP group range announcements from the Auto-RP packets that are denied by the boundary access list are removed. An Auto-RP group range announcement is permitted and passed by the boundary only if all addresses in the Auto-RP group range are permitted by the boundary access list. If any address is not permitted, the entire group range is filtered and removed from the Auto-RP message before the Auto-RP message is forwarded.

There is no similar command available on the Cisco IOS XR Software router.

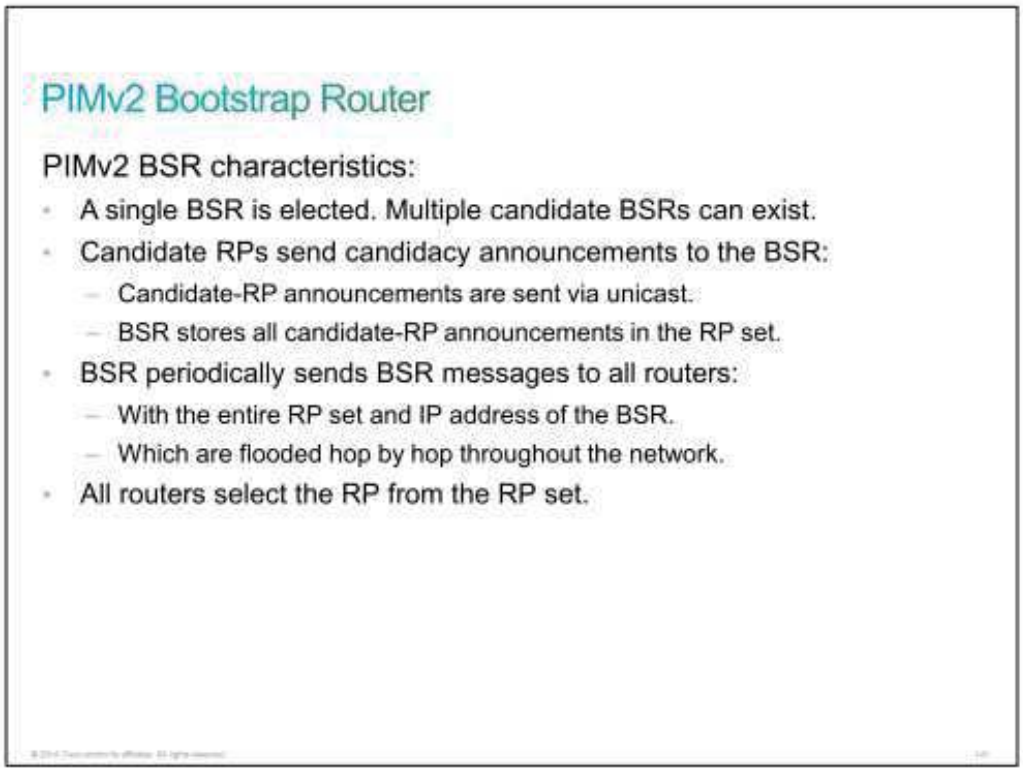
The administrative scoping is important because Auto-RP scoping allows any change to the RP designation to be configured only on the routers that are the RPs and the mapping agents not on the leaf routers.



https://t.me/learningnets

# PIMv2 Bootstrap Router

The BSR is an IETF standards mechanism available in PIMv2, in which a single router is elected as the BSR from a collection of candidate BSRs. If the current BSR fails, a new election is triggered. The election mechanism is preemptive based on the priority of the candidate BSR.




Via unicast, candidate RPs send candidate-RP announcements directly to the BSR. Candidate RPs learn the IP address of the BSR via periodic BSR messages.

The BSR does not elect the best RP for every group range it learns about. For every group range known, the BSR builds a set of candidate RPs, including all the routers that advertised their willingness to serve as the RP for the group range. The BSR stores the collection of candidate-RP announcements in a database that is called the RP set. The BSR periodically sends out BSR messages to the routers in the network to let them know that the BSR is still alive. BSR messages are flooded hop by hop throughout the network as multicasts to the all-PIM-routers group (224.0.0.13) with a TTL of 1. When a router receives a BSR message, it applies a RPF check, based on the source IP address in the packet. If the RPF check succeeds, the message is flooded out to all PIM-enabled interfaces.

BSR messages contain these elements:

- The RP set, consisting of candidate-RP announcements.
- The IP address of the BSR, so that candidate RPs know where to send their announcements.

Because the packets are flooded throughout the network, the routers will receive the BSR messages. Each receiving router selects the active RP for each group range using a common hash algorithm that is run against the RP set. The routers in the network will select the same RP for a given group range.

 © 2014 Cisco Systems, Inc.

https://t.me/learningnets

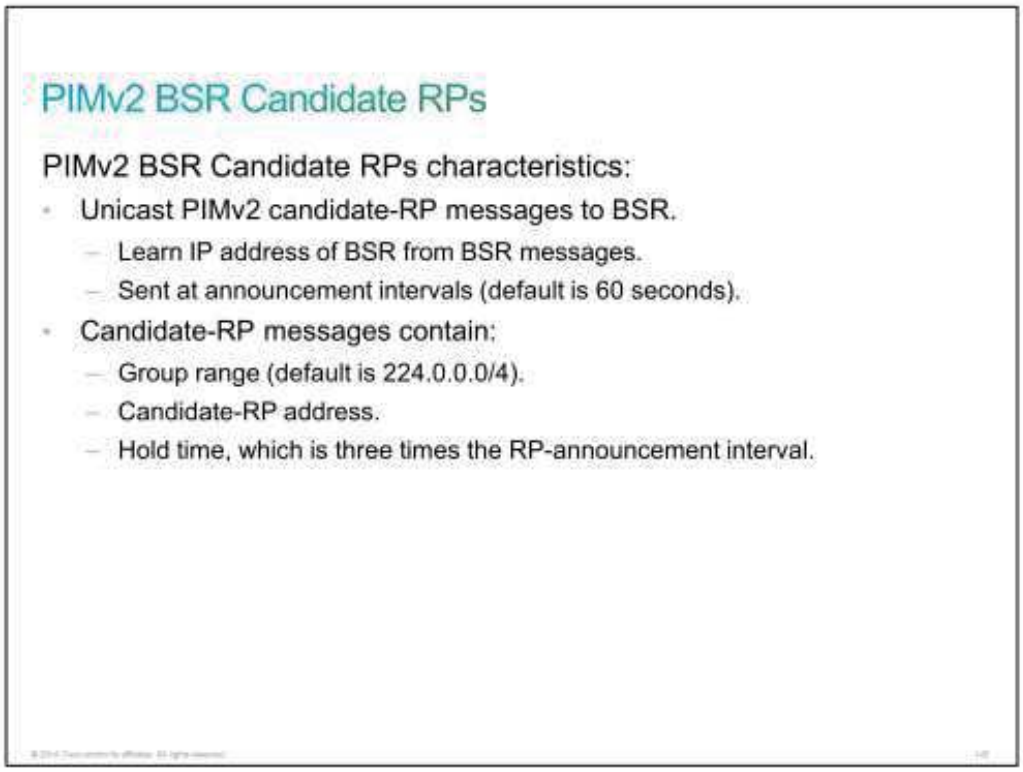
Contrary to the previously discussed Auto-RP method, administrative scoping cannot be used for BSR. Administrative scoping was not considered when BSR was designed. Candidate-RP announcements that are unicast to the BSR may cross multicast boundaries.



https://t.me/learningnets

# PIMv2 BSR Candidate RPs

Candidate RPs periodically send candidate-RP messages directly to the BSR via unicast. They know the BSR address because the BSR messages have been periodically flooded to the all-PIM-routers group (224.0.0.13). The default interval for the candidate-RP messages is 60 seconds.



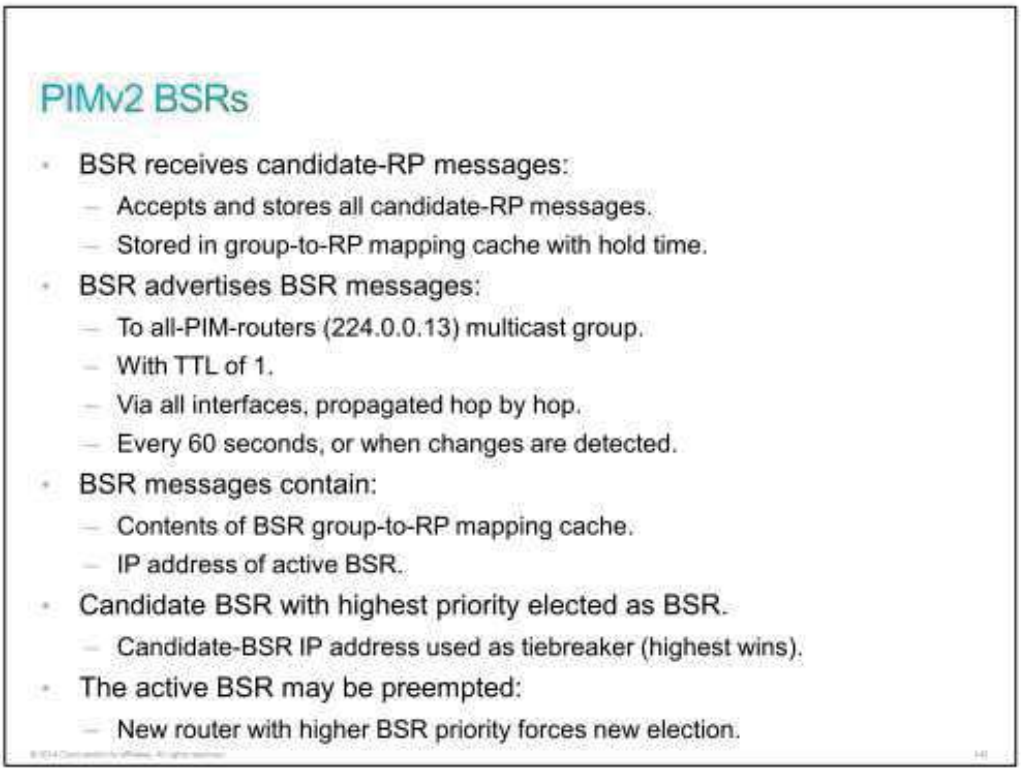

The candidate-RP announcement messages contain the group range, candidate-RP address, and a hold time.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# PIMv2 BSRs

The primary purpose of the BSR is to collect the candidate-RP announcements. These announcements are stored in a database called the RP set. The candidate-RP announcements are periodically flooded out to the other routers in the network using the BSR messages.




These are the attributes of a BSR message:

- It is sent periodically (the default is 60 seconds) by the BSR out the multicast interfaces.
- It is multicast to the all-PIM-routers group (224.0.0.13) with a TTL of 1. The PIM neighbors will receive these messages and retransmit them (again with a TTL of 1) out the interfaces except the one in which the message was received. An RPF check is done to ensure that the BSR message came in on the correct interface in the direction of the BSR.
- It contains the RP set and the IP address of the currently active BSR. This is how candidate RPs know where to unicast their candidate-RP messages.

Candidate BSRs participate in the BSR election mechanism using these rules:

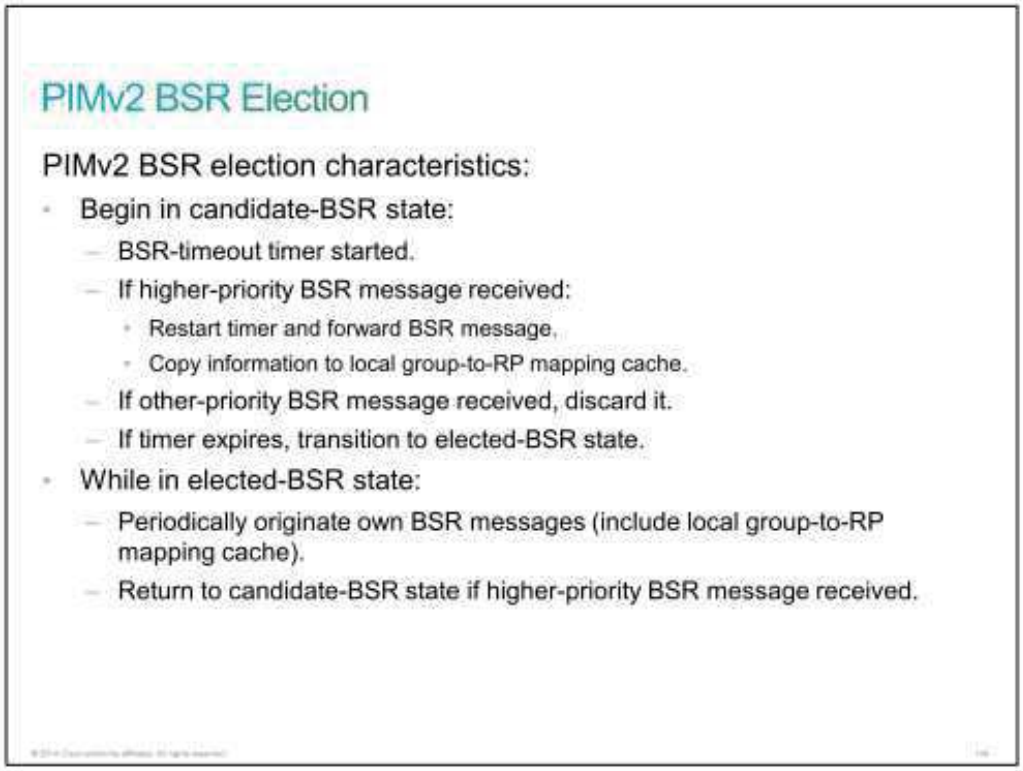
- The candidate BSR with the highest priority is elected as the BSR.
- The highest IP address of the candidate BSRs is used as a tiebreaker.
- The election mechanism is preemptive. If a new candidate BSR with a higher priority comes up, it triggers a new election process.



https://t.me/learningnets

# PIMv2 BSR Election

How and when routers in the network forward BSR messages plays a key role in the BSR election mechanism. The algorithm that is used to decide whether to process and forward an incoming BSR message depends on whether the router is a candidate BSR or not.




Candidate BSRs operate in one of two states: candidate-BSR state or elected-BSR state. Initially, a candidate BSR comes up in a candidate-BSR state.

In a candidate-BSR state:

- A BSR-timeout timer starts. If this timer expires, the router transitions to the elected-BSR state.
- If a BSR message is received with a higher priority than the priority of the candidate BSR, then the BSR (whose address is in the BSR message) is considered to be preferred. The BSR message is then processed as follows:
  - The BSR-timeout timer is reset.
  - The BSR message is forwarded out the other interfaces.
  - The RP set in the BSR message is copied into the local group-to-RP mapping cache.
- If a BSR message is received with a priority that is less than the priority of the candidate BSR, the BSR message is simply discarded.

In an elected-BSR state:

- The router is elected as the BSR and periodically originates BSR messages containing the current RP set.



https://t.me/learningnets

- If a BSR message is received from another router with a higher priority, the message is forwarded, and the router transitions back to a candidate-BSR state. Otherwise, the BSR message is discarded.




Routers that are not candidate BSRs operate in two states: accept-any state, and accept-preferred state. When a router that is not a candidate BSR boots, it starts in the accept-any state.

In the accept-any state, routers that are not the candidate BSRs perform as follows:

- Accept the first BSR message received.
- Copy the RP set into the local group-to-RP mapping cache.
- Save the current BSR priority and BSR address in the BSR message.
- Transition to the accept-preferred state.

In the accept-preferred state, routers that are not candidate BSRs perform as follows:

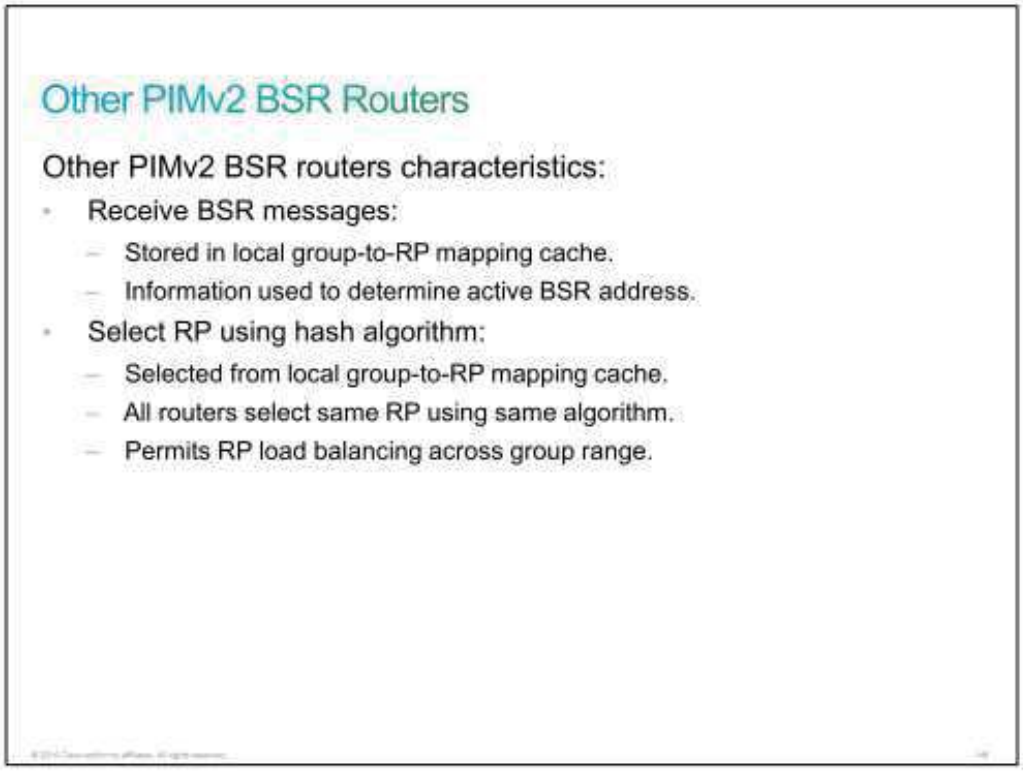
- Start the BSR-timeout timer with a period of 150 seconds. If this timer expires, the routers transition back to the accept-any state.
- Accept only BSR messages that are preferred. A preferred BSR message is one with a priority that is greater than or equal to the current BSR priority. The accepted BSR message is then processed as follows:
  - The BSR-timeout timer is reset. The BSR message is forwarded out the other interfaces except the one on which the message was received.
  - The RP set in the BSR message is copied into the local group-to-RP mapping cache.
  - The current BSR priority and the BSR address are saved in the BSR message.
  - If a BSR message is received with a priority that is less than that of the current BSR, the BSR message is simply discarded.



https://t.me/learningnets

# Other PIMv2 BSR Routers

The PIMv2 routers that are not BSRs or candidate RPs are described below:




PIMv2 routers perform as follows:

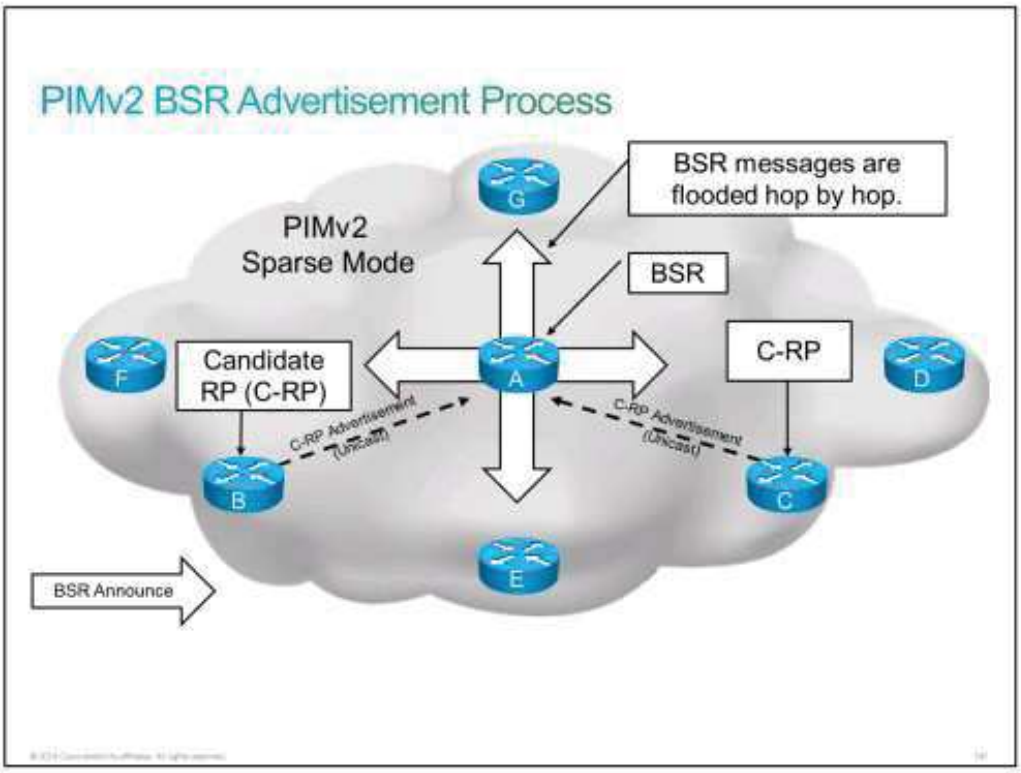
- Accept BSR messages that are based on the rules that are described earlier in this lesson. When a BSR message is accepted:
  - The RP set in the BSR message is stored in the local group-to-RP mapping cache.
  - The BSR message is forwarded out the other interfaces, except the one on which it was received.
- Select an RP using a hash algorithm:
  - The RP for a group is selected from the set of candidate RPs that have advertised their candidacy for a matching group range.
  - The routers use the same hashing algorithm to select the RP from the set of candidate RPs in the RP set. Since the routers run the same algorithm on the same RP set, they will select the same RP for a given group.
  - The hashing algorithm permits multiple candidate RPs to load-balance the duties of the RP across a range of groups. Only one candidate RP will be selected as the RP for any single group in the group range. However, the hash algorithm may select other candidate RPs as the RP for another group within the group range.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# PIMv2 BSR Advertisement Process

The PIMv2 BSR advertisement process is described below.




PIMv2 BSR advertisements process occurs as follows:

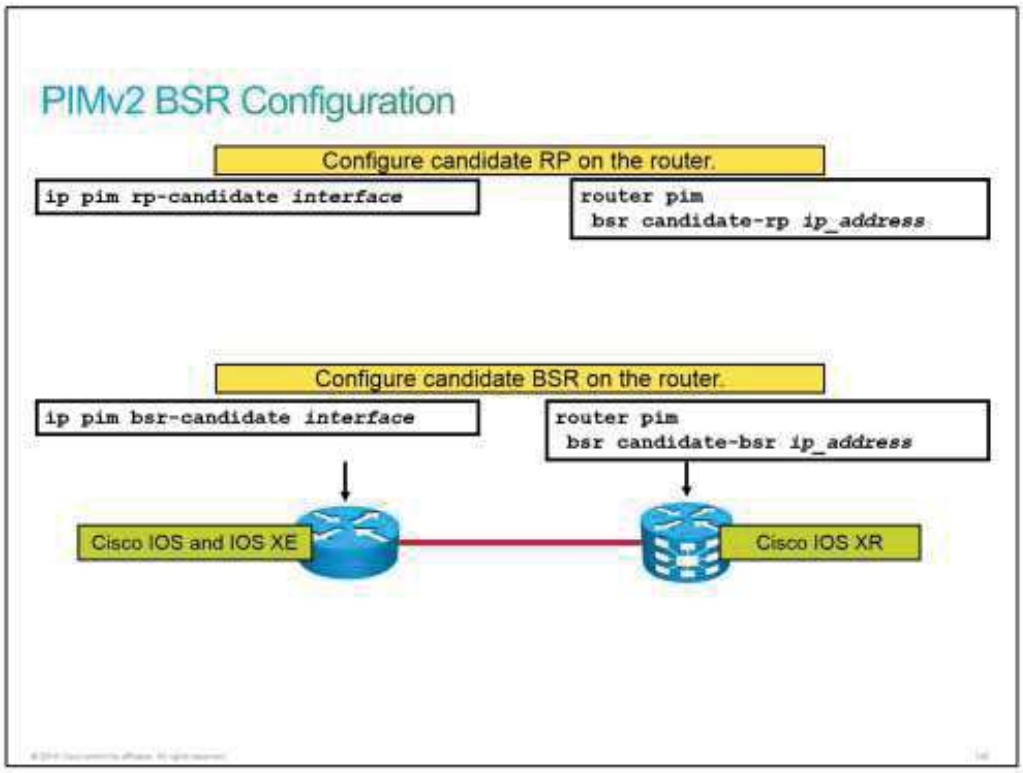
- Candidate RPs unicast their candidacy messages to the previously elected BSR. The candidate RPs learned the IP address of the BSR from the BSR messages that are being flooded throughout the network.
- The BSR receives and stores the candidate-RP information in a database that is called the RP set.
- The BSR periodically sends BSR messages containing the RP set out through its interfaces. These BSR messages are forwarded hop by hop away from the BSR by the routers in the network. Using a common hash algorithm, the routers will use the RP set to calculate the RP for a group.



https://t.me/learningnets

# PIMv2 BSR Configuration

The BSR configuration is described below:



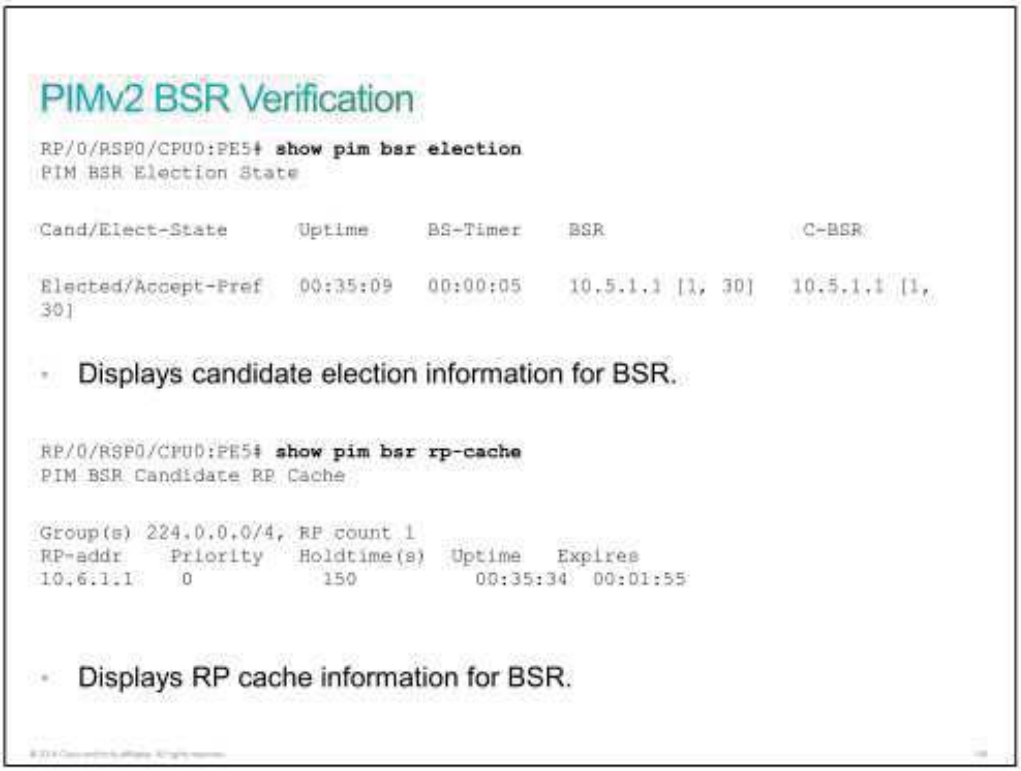

Candidate RPs are configured using the Cisco IOS and IOS XE **ip pim rp-candidate** command or the Cisco IOS XR **bsr candidate-rp** router PIM command. Candidate BSRs are configured using the Cisco IOS and IOS XE **ip pim bsr-candidate** command or the Cisco IOS XR **bsr candidate-bsr** router PIM command.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# PIMv2 BSR Verification

You can use the following commands to verify BSR operations. Use the **show pim bsr election** command to display BSR candidates election.



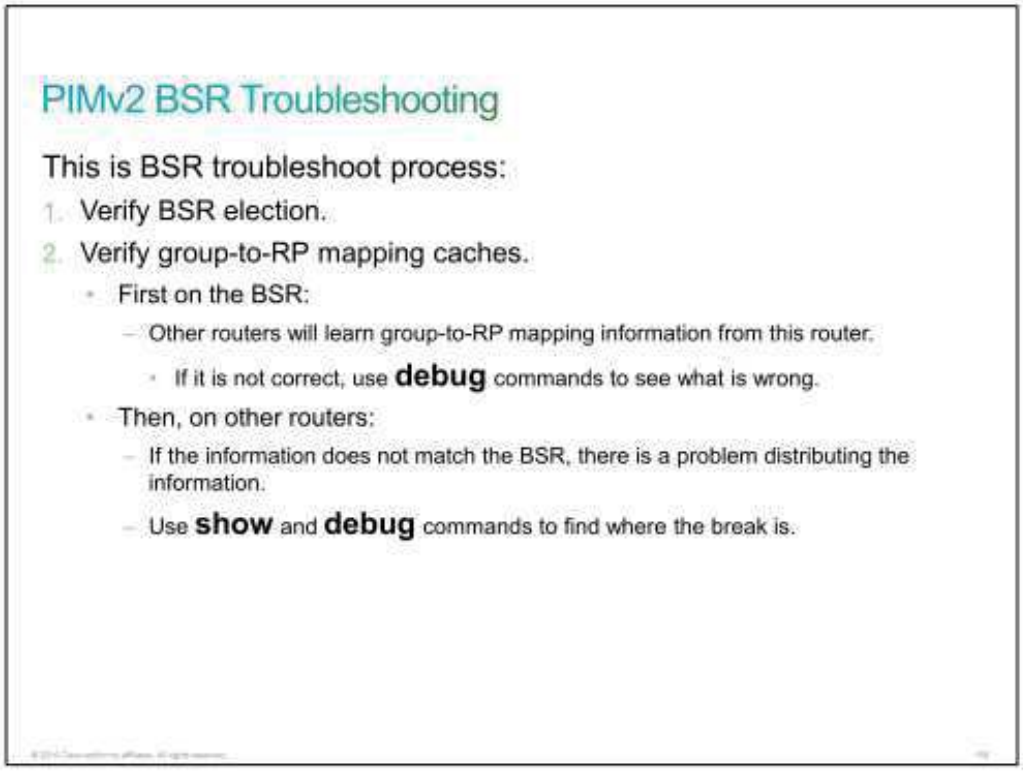

Use the **show pim bsr rp-cache** command to display RPs for specific multicast groups. In the example, the 10.6.1.1 router is the RP for all multicast groups that fall within 224.0.0.0/4.



https://t.me/learningnets

# PIMv2 BSR Troubleshooting

The BSR troubleshooting steps are described below:




Debug the BSR operation by verifying the election of the BSR router. Then, check the group-to-RP mapping caches on the elected BSR and other routers.

Because other routers in the network will learn the group-to-RP mapping information from the elected BSR, it is important that this information be correct on this router. If the information is not correct, you must verify that the PIMv2 candidate RPs are configured correctly. The BSR should receive the candidate-RP announcements.

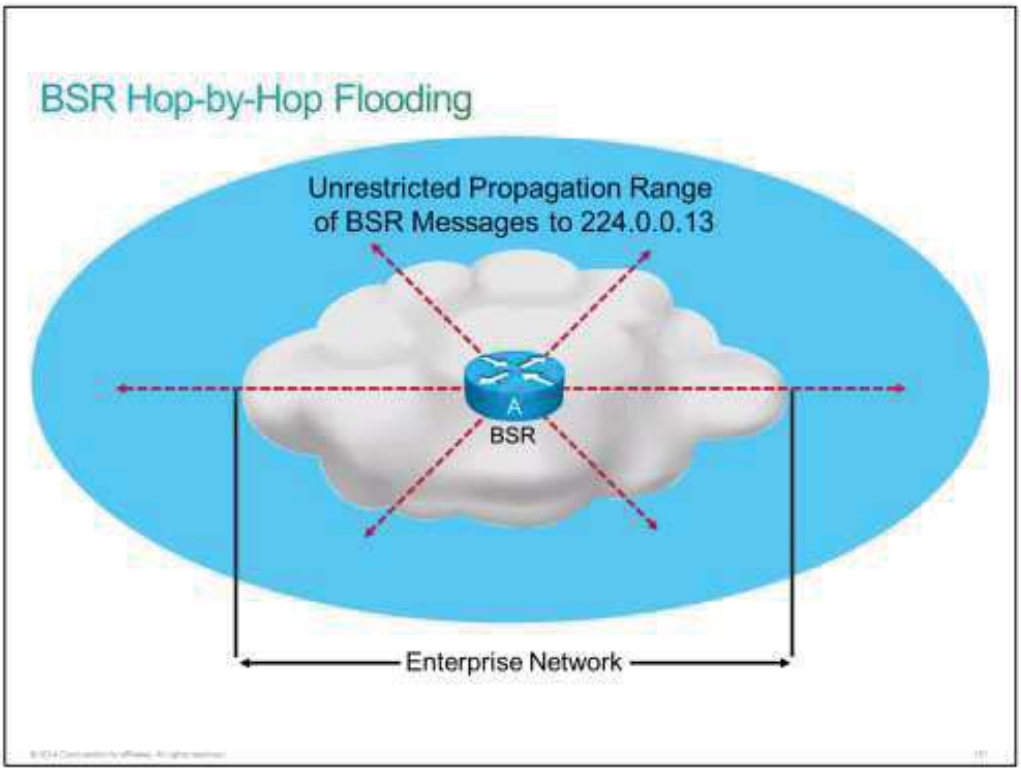
Verify the group-to-RP mapping caches on all other routers. Group-to-RP mapping information should match the BSR. If not, verify that the router is properly receiving BSR messages.



https://t.me/learningnets

# BSR Hop-by-Hop Flooding

BSR messages that are addressed to the multicast group 224.0.0.13 are propagated in a hop-by-hop fashion throughout the entire PIMv2 network. The leaf routers receive the packet, populate the local cache, and forward it across all other PIMv2 interfaces.



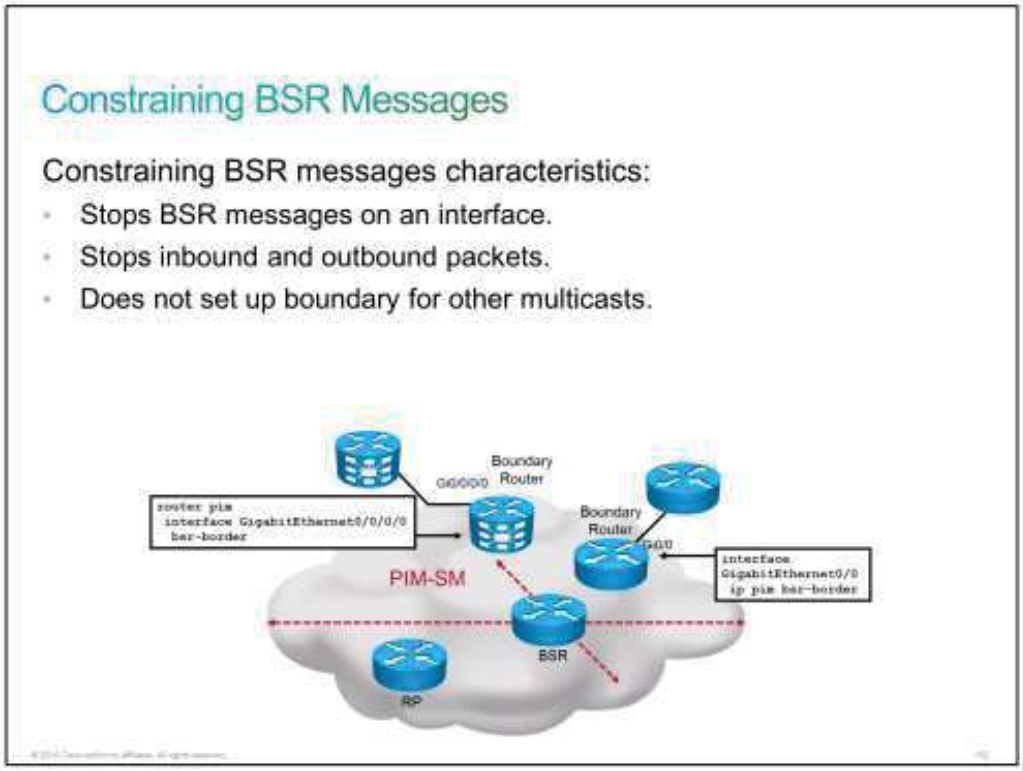

The problem arises on domain borders that should not leak the information about enterprise RPs. Leaking this information could affect network performance, as external domain RPs may be selected for the multicast distribution trees.



https://t.me/learningnets

# Constraining BSR Messages

Constrain BSR messages is described below:




You can set up an administrative boundary for BSR messages on an interface using the Cisco IOS and IOS XE **ip pim bsr-border** interface command or the Cisco IOS XR **bsr-border** router PIM command. When this command is configured on an interface, no PIMv2 BSR messages are sent or received through the interface. Using this command, you can configure an interface bordering another PIM domain to prevent BSR messages from being exchanged between the two domains. BSR messages should not be exchanged between different domains, because routers in one domain may elect RPs in the other domain. This will result in protocol malfunction or loss of isolation between the domains.

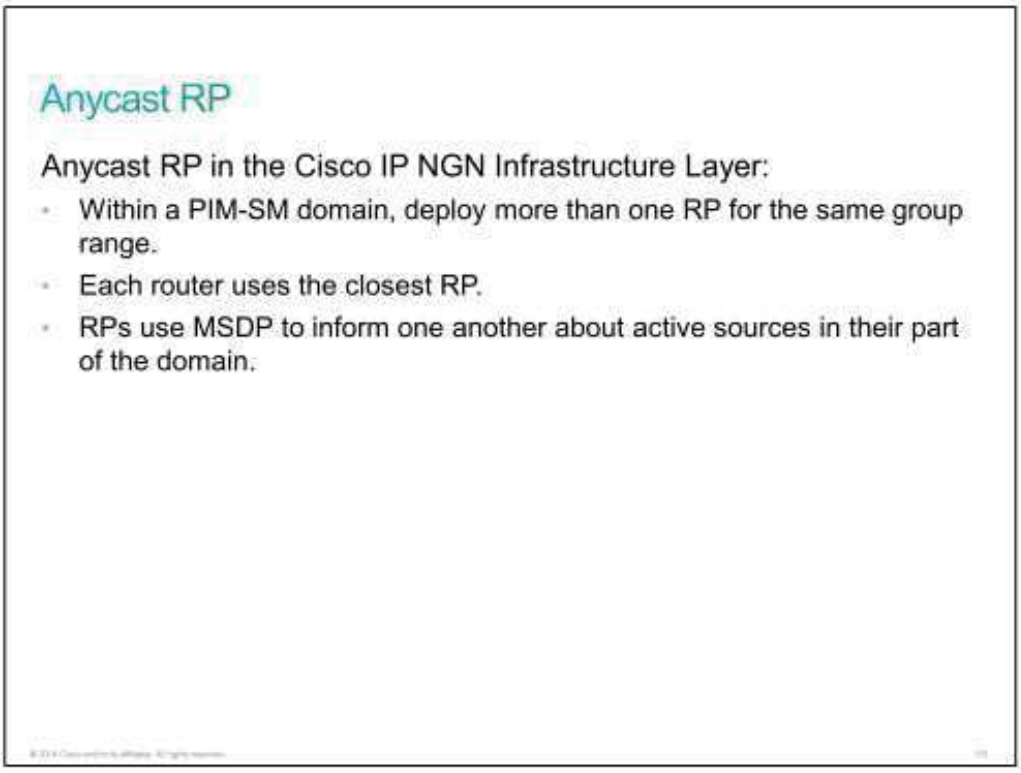
This command does not set up multicast boundaries, only a PIM domain BSR message border.



https://t.me/learningnets

# Anycast RP

Anycast RP works in concert with MSDP to provide RP redundancy, rapid RP failover, and RP load balancing. This concept is defined in RFC 3446.

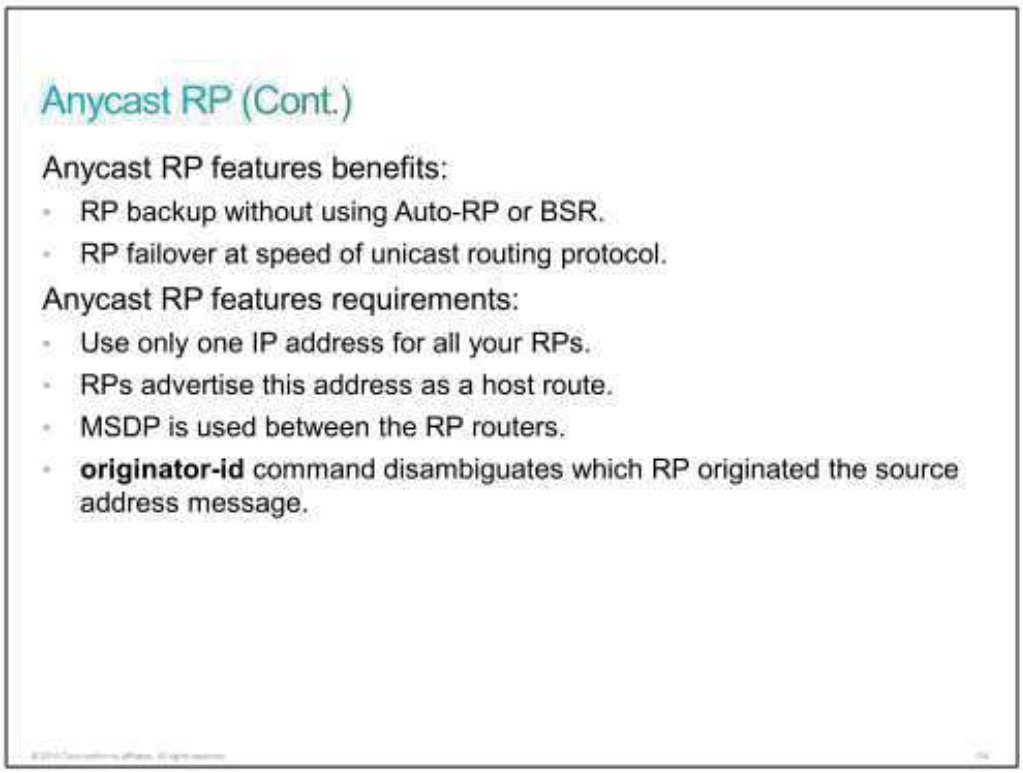



The Anycast RP mechanism works as follows:

- Two or more routers are configured as the active RP for the same group range at the same time. This is normally a configuration error that would partition the PIM-SM domain. However, MSDP is used to prevent this from happening.
- Each RP is assigned the same RP address. This is usually accomplished using a loopback interface and private address space. Each router advertises its RP address as a host route in the unicast routing protocol.
- Sources and receivers use the closest RP, based on their unicast routing table.
- The Anycast RPs also establish MSDP peerings. This allows each RP to learn which sources have been registered with the other Anycast RPs in the domain.
- The normal PIM-SM RP behavior causes the RP to join the source tree of active sources in the other parts of the network as necessary.



https://t.me/learningnets




The Anycast RP mechanism has these benefits:

- It enables RPs to provide for backup rendezvous points without having to use Auto-RP or BSR.
- RP failover occurs roughly at the same speed as the convergence of unicast routing protocol.

These are the requirements for Anycast RP:

- Anycast RPs are configured to use the same IP address.
- Anycast RPs advertise this IP address as a host route. This causes the other routers in the network to see only the closest RP.

If there is a metric tie, the normal RPF mechanism will select the path back to the RP. The path that is selected will be the one that has the highest next-hop address. Anycast RPs are tied together via MSDP peering sessions.

Anycast RPs provide for backup RPs without having to use Auto-RP or BSR. RP failover occurs at approximately the same speed as the unicast routing protocol converges.

All Anycast RPs are configured with the same RP IP address and should advertise this IP address as a host route. This causes the DR in the network to see only the closest RP.

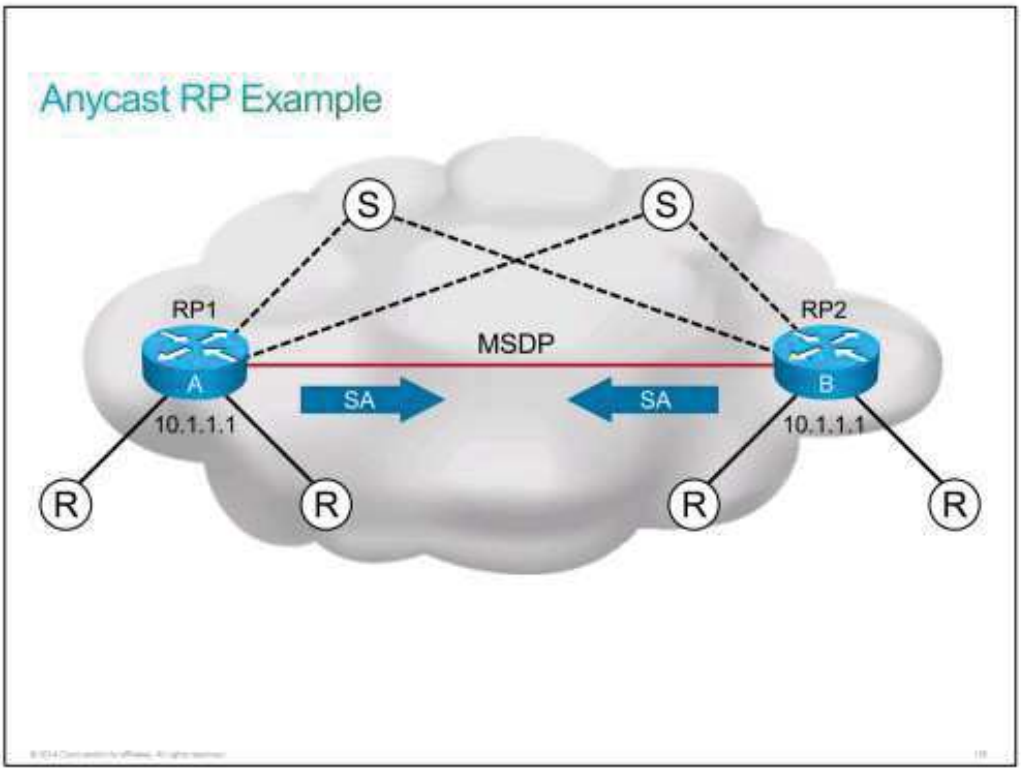
All Anycast RPs are tied together via MSDP peering sessions. The Cisco IOS and IOS XE **ip msdp originator-id** command or the Cisco IOS XR **originator-id** router MSDP command is used to control the IP address that is sent in SA messages. This precaution is taken to disambiguate which RP originated the SA message. If this action were not taken, all RPs would originate SA messages using the same IP address.



https://t.me/learningnets

# Anycast RP Example

In the figure, two Anycast RPs, RP1 and RP2, are configured with the same IP address (10.1.1.1). They also establish MSDP peering between them. To establish MSDP peering, these RPs must use some other address than 10.1.1.1.

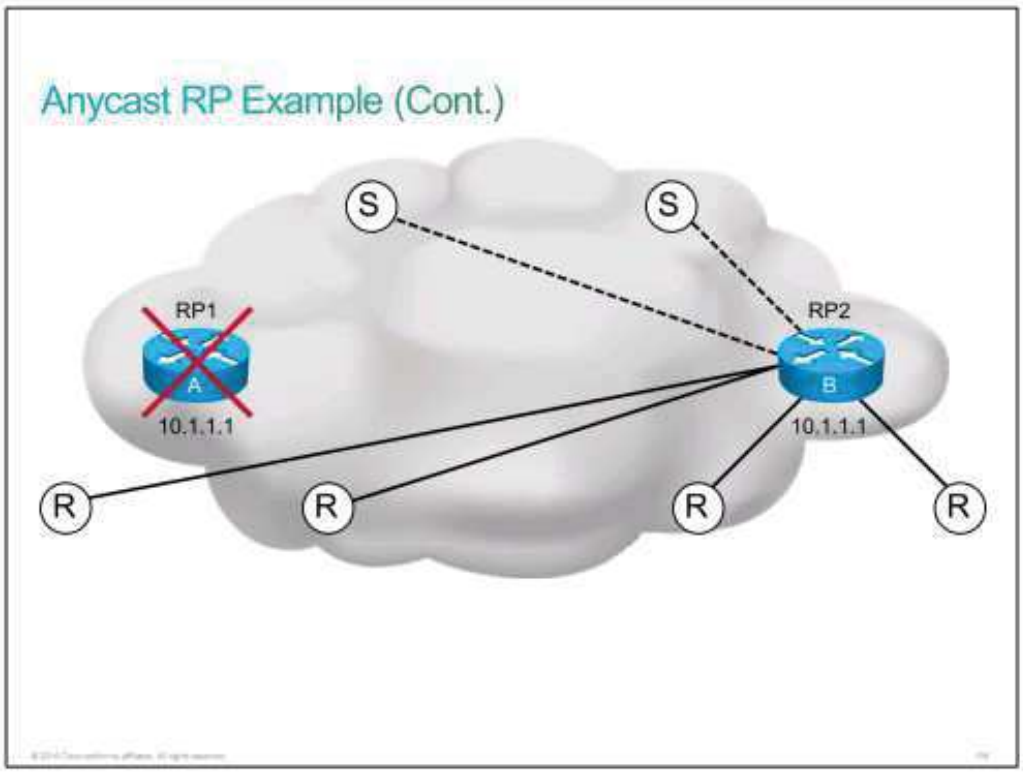



Initially, the DRs (not shown in the diagram) for the sources and receivers register to the closest RP based on their unicast routing table entry for IP address 10.1.1.1. This action causes the DRs in the left part of the network to register or join to RP1 while the DRs in the right part register or join to RP2.

When a new source registers with the nearest RP, that RP will send an MSDP SA message to its peer. This will cause the peer RP to join the SPT to the new source. Then, the peer RP can pull the source traffic to itself and send it down the shared tree to its receivers.



https://t.me/learningnets



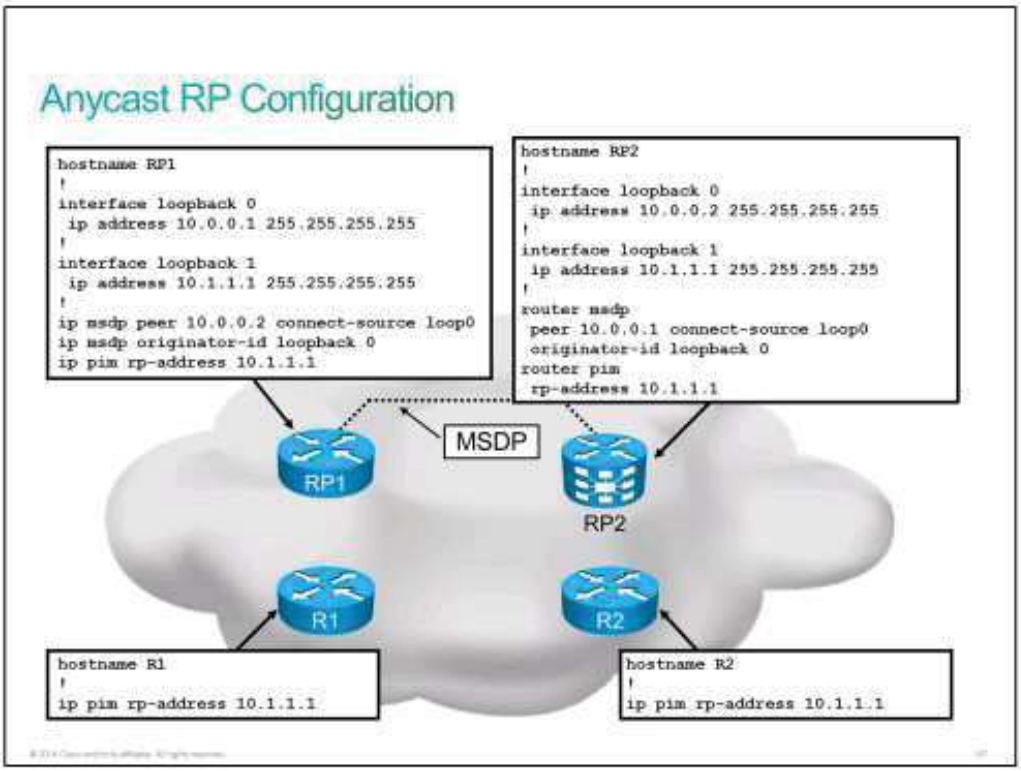

In this figure, RP1 goes down. When the unicast routing protocol reconverges, all the DRs (not shown in the figure) in the left part of the network can now see the route to 10.1.1.1 pointing toward RP2. The DRs in the left part of the network send new registers and joins to RP2, reestablishing the flow of traffic.



https://t.me/learningnets

# Anycast RP Configuration

In this configuration example, two Anycast RPs are configured with the same IP address, 10.1.1.1, using loopback 1.



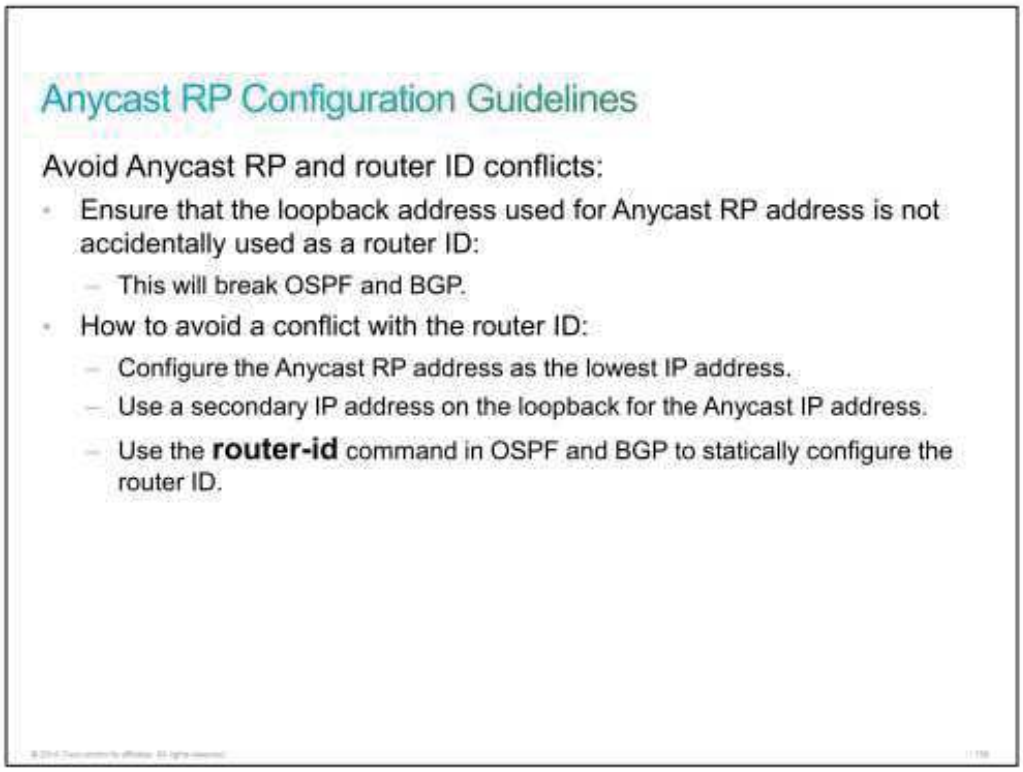

These RPs establish the MSDP sessions using the loopback 0 addresses, 10.0.0.1 and 10.0.0.2.



https://t.me/learningnets

# Anycast RP Configuration Guidelines

Care must be taken to prevent the loopback addresses that are being used for the Anycast RPs from being accidentally used as the router ID for OSPF and BGP. If this occurs, there will be multiple OSPF and BGP routers in the network with the same router ID, which can break the connectivity.




To avoiding router ID conflict, use one of these solutions:

- Configure the Anycast RP loopback address using the lowest IP address in the box.
- Configure a secondary address on the loopback address, and use this address for Anycast RP configuration.
- Use the Cisco IOS, IOS XE, and IOS XR **router-id** command to statically configure the OSPF and BGP router ID.



https://t.me/learningnets

# Summary

This topic summarizes the key points that were discussed in this lesson.

## Summary

- Static RP offers no fault tolerance.
- There are two protocols for automatic RP discovery: auto-RP and BSD.
- RP can be anywhere in the network, but smart placing can benefit scalability.
- With auto-RP, all routers learn RP address.
- Candidate RPs will advertise themselves.
- Mapping angents act as relays.
- All other routers will receive and forward auto-RP messages.
- To configure auto-RP, configure candidates and mapping agents.
- Troubleshooting auto-RP can be done in steps.
- Reach of auto-RP can be to great or too smal.
- Boundaries can be put into place to restrict the scope of auto-RP.
- BSR periodically sends relevant RP information to all routers.
- There is only one BSR, but there may be more BSR candidates.

## Summary (Cont.)

- BSR receives and forwards all candidate RP information.
- BSR is elected based on priority.
- If a router is neither BSR nor candidate BSR it will still process BSR messages.
- BSR messages are forwarded hop by hop.
- BSR is configured under PIM.
- When troubleshooting BSR, check election process and group-to-RP mapping cache.
- BSR messages may be flooded outside of safe bounds.
- Boundary can be configured to restrain BSR messages.
- Anycast RP can be used to provide RP redundancy without BSR or auto-RP.
- Anycast RP can coexist with MSDP.
- Anycast addresses must be configured for Anycast RP.
- Avoid conflicts when configuring anycast.

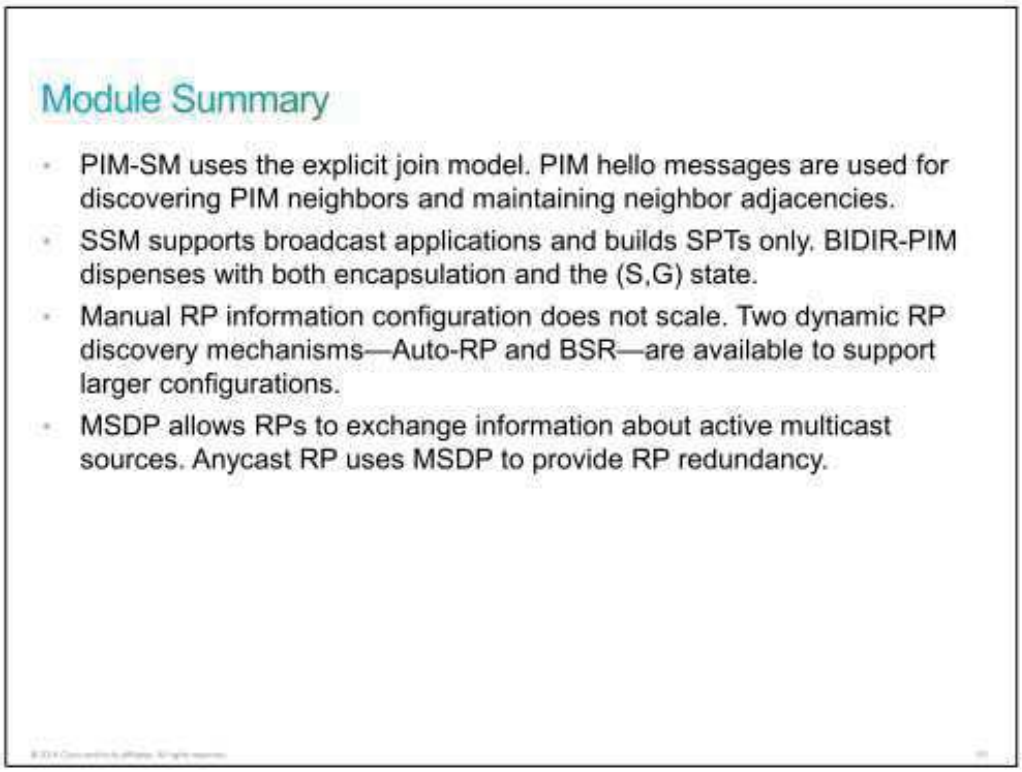


https://t.me/learningnets

https://t.me/learningnets

# Lesson 5: Module Summary

This topic summarizes the key points that were discussed in this module.




PIM-SM uses the explicit join model whereby receivers send PIM join messages to a designated RP. Manual RP information configuration does not scale to large environments. Two mechanisms of dynamic RP discovery have been developed to improve scalability in large environments: Auto-RP and BSR.

SSM is a datagram delivery model that best supports one-to-many applications. SSM uses all the benefits of sparse-mode protocols, but eliminates shared trees and only builds a SPT. The prerequisite for SSM deployment is IGMPv3.

BIDIR-PIM dispenses with both encapsulation and the (S,G) state. The major modification of PIM-SM to support bidirectional mode is the addition of a DF.



https://t.me/learningnets

For Auto-RP, you must configure one or more Auto-RP mapping agents and RPs. The two Auto-RP groups (224.0.1.39 and 224.0.1.40) operate in dense mode. Once the routers in the network learn the address of the elected RP, they will operate all other groups in sparse mode.

Via unicast, candidate RPs periodically send candidate-RP messages directly to the BSR. They know the BSR address, because the BSR periodically floods messages to the all-PIM-routers group (224.0.0.13).

MSDP has been designed for interdomain usage. It allows RPs to exchange information about active multicast sources. The RPs respond by originating (S,G) SA messages for active sources and sending them to their MSDP peers. Anycast RP uses MSDP within a PIM-SM domain to provide RP redundancy.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# Lesson 6: Module Self-Check

Use the questions here to review what you learned in this module. The correct answers and solutions are found in the Module Self-Check Answer Key.

1. An RPF check depends on which two tree types? (Choose two.) (Source: Introducing PIM-SM Protocol)

   A. Shared tree, uses RP address
   B. Source tree, uses source address
   C. Source tree, uses RP address
   D. Shared tree, uses source address

2. For which two is a PIM hello message used? (Choose two.) (Source: Introducing PIM-SM Protocol)

   A. Registering multicast sources
   B. Discovering PIM neighbors
   C. Joining and pruning multicast traffic
   D. Maintaining neighbor adjacencies

3. Which device sends a PIM register message? (Source: Introducing PIM-SM Protocol)

   A. Last-hop router
   B. First-hop router
   C. Rendezvous point
   D. Source
   E. Receiver



https://t.me/learningnets

4. Which multicast address does a PIM hello message use? (Source: Introducing PIM-SM Protocol)

   A. All-routers (224.0.0.2)
   B. All-routers (224.0.0.13)
   C. All-PIM-routers (224.0.0.13)
   D. All-PIM-routers (224.0.0.1)

5. In which three of these situations is an interface placed into the (*,G) OIL? (Choose three.) (Source: Introducing PIM-SM Protocol)

   A. A router has received a PIM join via this interface.
   B. A host has joined the group via IGMP.
   C. A router interface has been manually configured to join the group.
   D. A router has received a PIM prune via this interface.
   E. A host has left the group via IGMP.

6. Which three results happen if a router receives a join message and there is no (*,G) entry? (Choose three.) (Source: Introducing PIM-SM Protocol)

   A. The (*,G) join is ignored.
   B. A (*,G) entry is created.
   C. The interface is placed in the OIL.
   D. The (*,G) join is sent toward the RP.
   E. A (S,G) entry is created.

7. When does the RP send a PIM (S,G) register-stop message? (Source: Introducing PIM-SM Protocol)

   A. When multicast traffic is received natively
   B. When a receiver sends a PIM prune message
   C. When a source sends a PIM join message
   D. When the first-hop router sends a PIM join message

8. What does the SPT threshold "infinity" mean? (Source: Introducing PIM-SM Protocol)

   A. Never join the shared tree.
   B. When the first packet arrives, join the SPT.
   C. Never join the SPT.
   D. The SPT threshold cannot be set to infinity.

9. When does a router send a PIM pruning message? (Source: Introducing PIM-SM Protocol)

   A. When a receiver sends an IGMP leave message
   B. When the OIL is empty
   C. When there is only one interface in the OIL
   D. When a receiver sends a PIM prune message

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

10. What is the duration of the (S,G) entry expiration timer? (Source: Introducing PIM-SM Protocol)

A. 1 second
B. 3 seconds
C. 30 seconds
D. 3 minutes

11. For which two situations is SSM suitable? (Choose two.) (Source: Implementing PIM-SM Enhancements)

A. For well-known sources
B. For broadcast applications
C. For random sources
D. For many-to-many applications (video conference)

12. For which two reasons is IGMPv3 a prerequisite for SSM deployment? (Choose two.) (Source: Implementing PIM-SM Enhancements)

A. The host can specify the RP address.
B. The host can specify the group.
C. The host can specify the source.
D. The source can specify the group and destination.

13. What is the default SSM range of IP multicast addresses? (Source: Implementing PIM-SM Enhancements)

A. 223.0.0.0/8
B. 232.1.0.0/24
C. 232.0.0.0/8
D. 232.1.0.0/16

14. Which Cisco IOS/IOS XE command is used to display information about SSM mapping? (Source: Implementing PIM-SM Enhancements)

A. **show ip igmp groups**
B. **show host**
C. **debug ip igmp group-address**
D. **show ip igmp ssm-mapping**

15. Which two conditions are benefits of BIDIR-PIM? (Choose two.) (Source: Implementing PIM-SM Enhancements)

A. All packets are encapsulated.
B. The first packets from the source are not encapsulated.
C. There are no (*,G) states.
D. There are no (S,G) states.



https://t.me/learningnets

16. Which router will become DF on the segment? (Source: Implementing PIM-SM Enhancements)

A. The router with the best route to the receivers
B. The router with the best route to the RP
C. The router with the best route to the source
D. The router with the best route to the BSR

17. Which condition would *not* trigger the process of re-electing a DF? (Source: Implementing PIM-SM Enhancements

A. A change in the unicast metric to reach the RP
B. A change in the interface on which the RP is reachable
C. Periodic triggering, every 3 minutes
D. A new PIM neighbor on a link
E. Death of an elected DF

18. What is the GLOP address range for AS 54321? (Source: Implementing Interdomain IP Multicast)

A. 233.212.196.0/24
B. 233.212.49.0/24
C. 233.196.212.0/24
D. 233.54.32.1/32

19. What is the SSM range of the multicast IP addresses? (Source: Implementing Interdomain IP Multicast)

A. 232.0.0.0/8
B. 233.0.0.0/8
C. 224.0.0.0/8
D. 223.0.0.0/8

20. Why is MSDP used in the IP multicast-enabled network? (Source: Implementing Interdomain IP Multicast)

A. To exchange multicast routing information
B. To announce sources in one domain to other domains
C. To exchange unicast routing information
D. To announce receivers in one domain to other domains

21. MSDP can work with both PIM-DM and PIM-SM. (Source: Implementing Interdomain IP Multicast)

A. True
B. False

22. Which mechanism prevents SA messages from looping through the Internet? (Source: Implementing Interdomain IP Multicast)

A. TTL value
B. Poison reverse
C. Count to infinity
D. RPF check

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

23. Which statement describes the peer that initiates the MSDP connection and the duration of the keepalive timer? (Source: Implementing Interdomain IP Multicast)

A. The peer with the lower IP address initiates the connection, and the keepalive timer is set at 75 seconds.
B. The peer with the higher IP address initiates the connection, and the keepalive timer is set at 75 seconds.
C. The peer with the lower IP address initiates the connection, and the keepalive timer is set at 60 seconds.
D. The peer with the higher IP address initiates the connection, and the keepalive timer is set at 60 seconds.

24. Which messages are used in MSDP? (Source: Implementing Interdomain IP Multicast)

A. Keepalive
B. SA
C. SA request
D. SA response
E. All of the above

25. What is the correct order of the RPF-check rules that apply when an SA message is received? (Source: Implementing Interdomain IP Multicast)

| | |
|---|---|
| Sending MSDP peer does not equal MP-BGP peer. | 2. |
| Sending MSDP peer equals MP-IBGP peer. | 1. |
| Sending MSDP peer equals MP-EBGP peer. | 3. |

26. Which three options describe benefits of using the MSDP SA caching mechanism? (Choose three.) (Source: Implementing Interdomain IP Multicast)

A. Reduces join latency
B. Acts as a valuable debugging tool
C. Helps prevent source-active storms
D. Brings receivers closer to the sources
E. Consumes less memory

27. What information is cached by the MSDP SA caching server? (Source: Implementing Interdomain IP Multicast)

A. All (S,G) pairs received by the source-active messages
B. Only (*,G) pairs received by the source-active messages
C. Only (S,G) pairs where receivers are listening
D. Only (S,G) pairs from directly connected domains



https://t.me/learningnets

28. The pseudo-MSDP peer concept is often used on an MP-BGP route reflector. (Source: Implementing Interdomain IP Multicast)

A. True
B. False

29. What is a necessary step in MSDP configuration? (Source: Implementing Interdomain IP Multicast)

A. Configuring an MSDP peer
B. Maintaining an MSDP connection
C. Controlling MSDP source-active message origination and propagation
D. Configuring MSDP source-active caching and requesting

30. Which two addresses are involved in Auto-RP communications? (Choose two.) (Source: Identifying Rendezvous Point Distribution Solutions)

A. 224.0.0.39
B. 224.0.1.39
C. 224.0.0.40
D. 224.0.1.40

31. Which two addresses are involved in BSR communications? (Choose two.) (Source: Identifying Rendezvous Point Distribution Solutions)

A. 224.0.0.1
B. 224.0.0.2
C. 224.0.0.13
D. Unicast to the BSR

32. Which method supports administrative scoping? (Source: Identifying Rendezvous Point Distribution Solutions)

A. Auto-RP
B. BSR
C. Auto-RP and BSR
D. None of the above

33. Which two configuration steps are mandatory for Auto-RP to operate? (Choose two.) (Source: Identifying Rendezvous Point Distribution Solutions)

A. Configuring the candidate RP
B. Configuring the BSR
C. Configuring the candidate BSR
D. Configuring the mapping agent

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

34. Which routers of an Auto-RP solution can perform RP information filtering? (Source: Identifying Rendezvous Point Distribution Solutions)

A. RP routers
B. Mapping agent routers
C. Leaf routers
D. RP and mapping agent routers

35. Which two criteria are required in the election of the BSR? (Choose two.) (Source: Identifying Rendezvous Point Distribution Solutions)

A. Candidate-BSR highest priority
B. Candidate-BSR highest IP address
C. Candidate-BSR lowest priority
D. Candidate-BSR lowest IP address

36. How do candidate RPs communicate with the BSR? (Source: Identifying Rendezvous Point Distribution Solutions)

A. Multicast IP 224.0.1.39
B. Multicast IP 224.0.1.40
C. Multicast IP 224.0.0.13
D. Unicast to the BSR IP address

37. What steps are involved in troubleshooting the BSR mechanism? (Source: Identifying Rendezvous Point Distribution Solutions)

A. Verify BSR election
B. Verify group-to-RP mapping caches on BSRs
C. Verify group-to-RP mapping caches on non-BSRs
D. All of the above

38. How can MSDP be used inside a single domain? (Source: Identifying Rendezvous Point Distribution Solutions)

A. Use full-mesh MSDP sessions between all routers.
B. Use MSDP instead of BGP.
C. Use MSDP sessions between RPs.
D. Use MSDP only between domains.

39. Which MSDP packets carry relevant MSDP information? (Source: Identifying Rendezvous Point Distribution Solutions)

A. SA messages
B. MSDP peer keepalive messages
C. MSDP (*,G) join messages
D. MSDP (S,G) join messages



https://t.me/learningnets

40. Is full mesh required in MSDP? (Source: Identifying Rendezvous Point Distribution Solutions)

A. No
B. Yes
C. Yes, interdomain only
D. Yes, intradomain only

41. What kind of distribution trees result from MSDP and Anycast RP? (Source: Identifying Rendezvous Point Distribution Solutions)

A. Source tree
B. Shared tree
C. Shared tree from source to local RP, and source tree from local RP to receiver
D. Source tree from source to local RP, and shared tree from local RP to receiver

42. What should be considered when deploying Anycast RP? (Source: Identifying Rendezvous Point Distribution Solutions)

A. MSDP sessions that follow BGP sessions
B. Enabling full-mesh MSDP sessions
C. OSPF metric
D. Router ID conflicts for OSPF and BGP

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# Module Self-Check Answers

## Answer Key

1. A, B
2. B, D
3. B
4. C
5. A, B, C
6. B, C, D
7. A
8. C
9. B
10. D
11. A, B
12. B, C
13. C
14. D
15. B, D
16. B
17. C
18. B
19. A
20. B
21. B
22. D
23. C



https://t.me/learningnets

24. E

25. Sending MSDP peer equals MP-EBGP peer. — 2.

    Sending MSDP peer equals MP-IBGP peer. — 1.

    Sending MSDP peer does not equal MP-BGP peer. — 3.

26. A, B, C

27. A

28. A

29. A

30. B, D

31. C, D

32. A

33. A, D

34. C

35. A, B

36. D

37. D

38. C

39. A

40. A

41. A

42. D

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# Module 6: Service Provider IPv6 Transition Implementations

## Introduction

Because of the near depletion of public IPv4 address space, service providers have to consider migration to IPv6. Several transition technologies are available, and each service provider has to determine which one suits it best. When migrating to IPv6, the service providers also should consider migration of services, such as DNS and DHCP, to IPv6.

This module first describes services that are available on IPv6. These include DNS, DHCPv6, IPv6 multicast, and IPv6 QoS. The module then discusses transition technologies that are available to migrate to IPv6. The module concludes with deployment strategies that the service providers are facing when deploying IPv6.

Upon completing this module, you will be able to:

- Describe DNS and DHCP support for IPv6 and support for QoS and multicast in the IPv6 network
- Describe IPv6 transition mechanisms and which methods are most effective in the service provider network
- Describe the deployment strategies that service providers are facing when deploying IPv6



https://t.me/learningnets

https://t.me/learningnets

# Lesson 1: Introducing IPv6 Services

## Overview

As service providers migrate addressing and routing protocols to IPv6, all other services should migrate to IPv6 as well before IPv6 connectivity could be offered to customers. Among others, these services include multicast, DNS, DHCP, QoS, and troubleshooting and management tools. It is an outstanding fact that many of the applications and protocols that support IPv4 support IPv6 as well and use the same principles. The most obvious change in configuration and operations is the change in addressing. However, there are services and applications that were either reimplemented or invented with IPv6. An example is how clients automatically obtain IPv6 addresses using stateless autoconfiguration.

The lesson first describes multicast for IPv6. The lesson continues with DNS and DHCP services for IPv6. The lesson then covers QoS support for IPv6 and concludes with troubleshooting and management tools that are available for IPv6 on Cisco routers.

Upon completing this lesson, you will be able to:

- Describe multicast in the Cisco IP NGN infrastructure layer
- Describe the IPv6 multicast address format
- Describe the IPv6 multicast address scope
- Describe the IPv6 solicited-node multicast address format
- Describe IPv6 multicast address with a global scope
- Compare IPv4 and IPv6 multicast
- Provide an overview of PIMv6
- Describe the embedding of the RP address in an IPv6 multicast address
- Discuss IPv6 multicast routing configuration
- Describe the MLD protocol



https://t.me/learningnets

- List the MLDv1 messages
- Describe the MLDv1 general query message
- Describe the MLDv1 report message
- Describe the MLDv1 done message
- Describe the MLDv1 address-specific query message
- Describe the MLDv2 protocol
- Describe MLD access groups and group limits
- Discuss MLD configuration
- Discuss MLD verification
- Discuss MLD join-group and static-group configuration options
- Discuss MLD snooping operations
- Describe DNS and DHCPv6 in the Cisco IP NGN infrastructure layer
- Describe DNS IPv6 support
- Describe dynamic DNS IPv6 support
- Describe DHCPv6 operations
- Describe DHCPv6 server router configuration
- Describe DHCPv6 Lite operations
- Describe DHCPv6 prefix delegation
- Describe DHCPv6 verification commands on the router
- Describe QoS in the Cisco IP NGN infrastructure layer
- Describe the fields that are used in the IPv6 header to support QoS functions
- Describe the IPv6 Traffic Class field
- Describe the IPv6 Flow Label field
- Describe IPv6 QoS configuration
- Describe the need for Cisco IOS Software features to support both IPv4 and IPv6
- Describe IPv6 Telnet and SSH server and client support in Cisco IOS Software
- Describe other IPv6 tools in Cisco IOS
- Describe Cisco Discovery Protocol support for IPv6
- Describe Cisco Express Forwarding support for IPv6
- Describe IP SLA support for IPv6
- Describe how to configure IP SLA

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# Multicast in the Cisco IP NGN Infrastructure Layer

Multicast is used in the infrastructure layer of the Cisco IP Next-Generation Network. Multicast is implemented on the service provider core and IP edge devices.

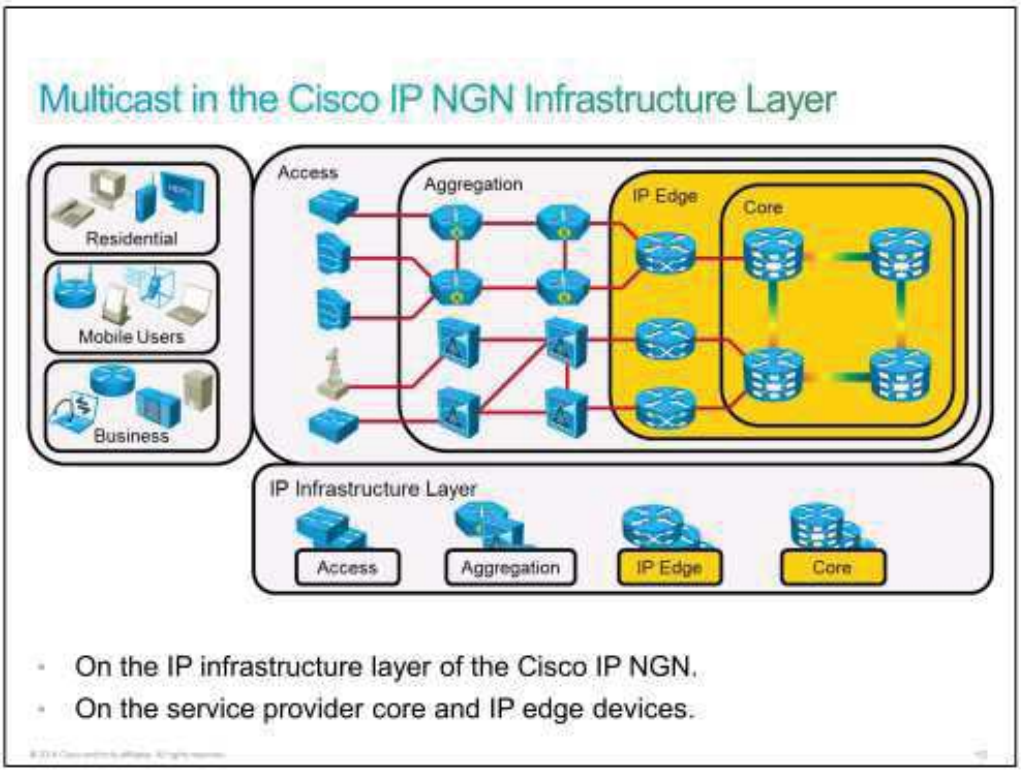


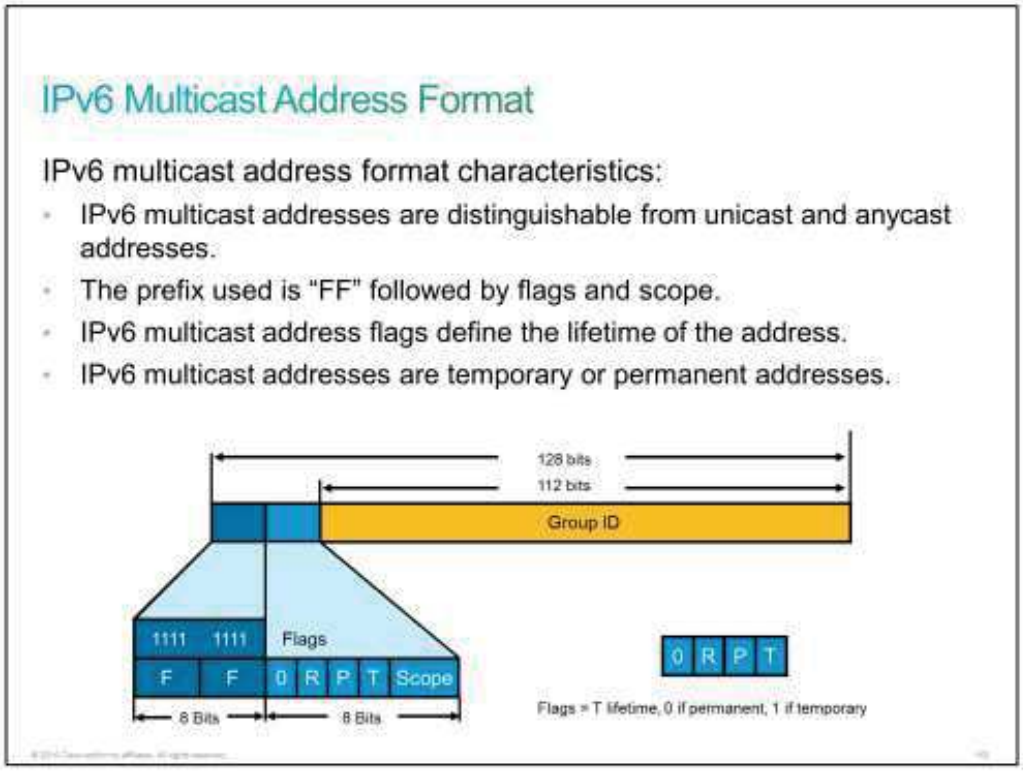

Figure shows where multicast is used in the Cisco IP NGN Infrastructure Layer.



https://t.me/learningnets

# IPv6 Multicast Address Format

The IPv6 multicast address format is described below:




An IPv6 multicast address is an IPv6 address that has a prefix of FF00::/8 (1111 1111). The FF at the start of the address identifies the address as being a multicast address.

An IPv6 multicast address is an identifier for a set of interfaces that typically belong to different nodes. A node may belong to any number of multicast groups. A packet that is sent to a multicast address is delivered to all interfaces that are identified by the multicast group address.

The second octet following the prefix defines the lifetime and scope of the multicast address.

Marks are a set of four highlights:

- The high-order (bit) mark is reserved and must be initialized to 0.
- T = 0 indicates a permanently assigned ("well-known") multicast address, which is assigned by the IANA.
- T = 1 indicates a nonpermanently assigned ("transient") multicast address.
- The P highlight's definition and usage can be found in RFC 3306.
- The R highlight's definition and usage can be found in RFC 3956.

  The P and R highlights are also explained later in this lesson.

The 112-bit Group ID identifies the multicast group, either permanent or transient, within the given scope.



https://t.me/learningnets

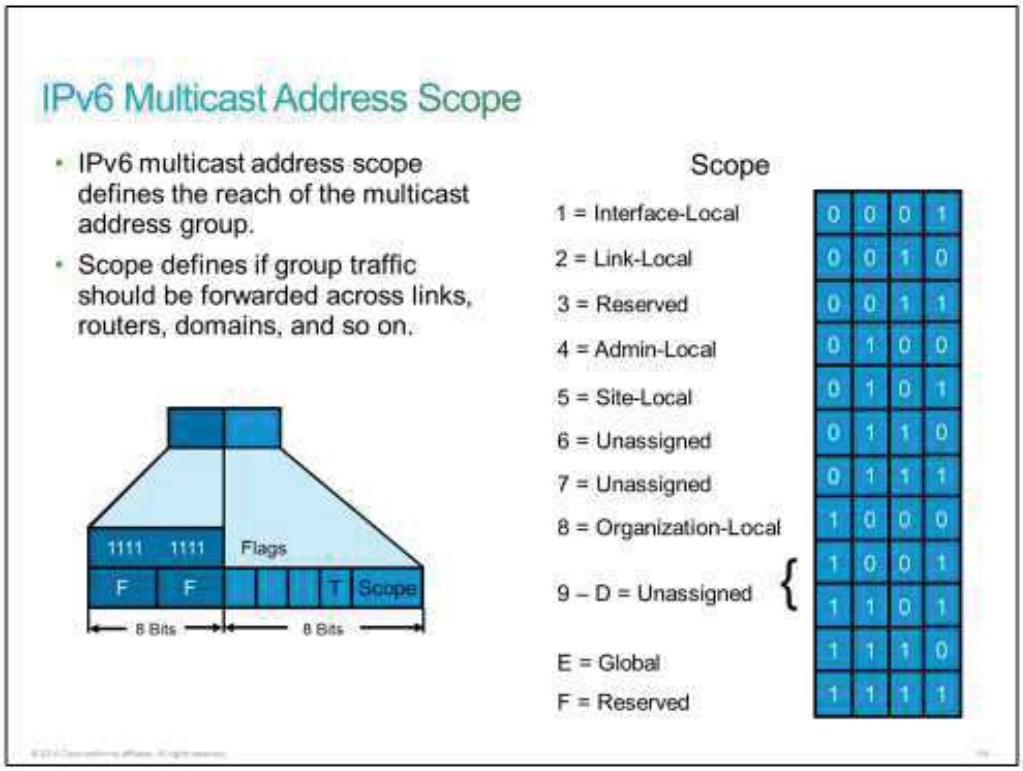
| | |
|---|---|
| **Note** | For IPv6 multicast addresses, the Ethernet MAC address is derived by the four low-order octets that are logically formed with the MAC address 33:33:00:00:00:00. For example, the IPv6 address FF02::1:3 would map to the Ethernet MAC address 33:33:00:00:01:03. |



https://t.me/learningnets

# IPv6 Multicast Address Scope

The IPv6 multicast address scope field uses a single, 4-bit field that is embedded in the IPv6 multicast address format.




The Scope field is used to limit the scope of the multicast group. There are several values that are defined, and others that are reserved or unassigned:

- 0 is reserved.
- 1 is interface-local scope.
- 2 is link-local scope.
- 3 is reserved.
- 4 is admin-local scope.
- 5 is site-local scope.
- 6 is unassigned.
- 7 is unassigned.
- 8 is organization-local scope.
- 9–D is unassigned.
- E is global scope.
- F is reserved.

Consequently, a multicast address that has the scope of an interface, link, site, organization, or a global scope has a scope parameter of 1, 2, 5, 8, or E, respectively.



https://t.me/learningnets

For example, a multicast address with the prefix FF02::/16 is a permanent multicast address with a link scope.

The scope values have these various functions:

- The interface-local scope spans only a single interface on a node and is useful only for loopback transmission of multicast. Packets with this destination address may not be sent over any network link but must remain within the current node; this is the multicast equivalent of the unicast loopback address.
- The link-local multicast scope spans the same topological region as the corresponding unicast scope. Packets with this destination address may not be routed anywhere.
- The admin-local scope is the smallest scope that must be administratively configured; for example, it is not automatically derived from physical connectivity or another, nonmulticast-related configuration.
- The site-local scope is intended to span a single site.
- The organization-local scope is intended to span multiple sites belonging to a single organization.
- Scopes labeled "unassigned" are available for administrators to define additional multicast regions.
- Packets with the global scope destination address are eligible to be routed over the public Internet.

Routers must not forward any multicast packets beyond the scope that is indicated by the Scope field in the destination multicast address.

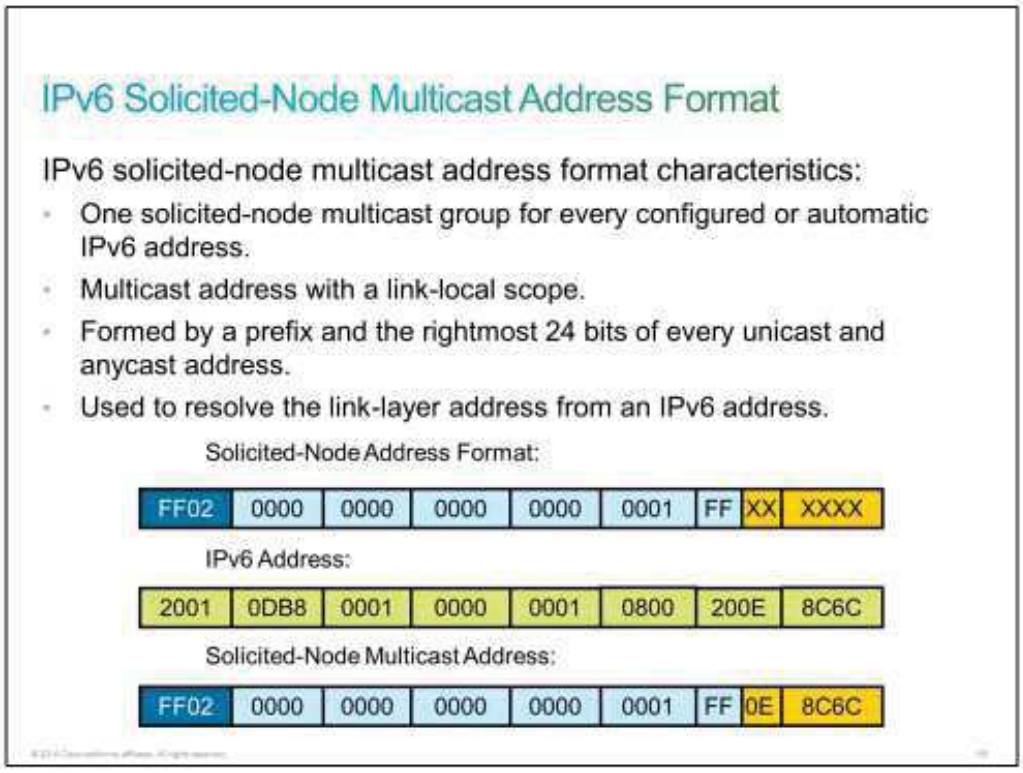
Nodes must not originate a packet to a multicast address whose scope field contains the reserved value 0. If such a packet is received, it must be silently dropped. Nodes should not originate a packet to a multicast address whose scope field contains the reserved value F. If such a packet is sent or received, it must be treated the same as packets that are destined to a global (scope E) multicast address.



https://t.me/learningnets

# IPv6 Solicited-Node Multicast Address Format

Solicited-node multicast addresses are used in NDP for obtaining the Layer 2 link-layer addresses of other nodes.




Solicited-node multicast addresses are used with IPv6 NDP to provide the same function as ARP in IPv4. ARP uses broadcasts to send an ARP request to the broadcast address 255.255.255.255, which is received by all stations on the local link, although only one station would need to respond. The other stations still have to process and discard the request. This interruption can cause problems on networks if the amount of broadcast traffic becomes excessive. Because a solicited-node multicast address is a function of the last 24 bits of an IPv6 unicast (or anycast) address, the number of hosts that are subscribed to each solicited-node multicast address is very small. This number typically would be one, but there could be a few because the mapping function is not a 1:1 mapping. This means that a host should not need to be interrupted as often to service neighbor solicitation requests, compared to ARP in IPv4.

The following outlines the format of the IPv6 solicited-node address:

FF02:0:0:0:0:1:FFXX:XXXX

Solicited-node multicast addresses are computed as a function of a node unicast and anycast addresses. A solicited-node multicast address is formed by taking the low-order 24 bits of an address (unicast or anycast) and appending those bits to the prefix FF02:0:0:0:0:1:FF00::/104 resulting in a multicast address in the range

FF02:0:0:0:0:1:FF00:0000

to

FF02:0:0:0:0:1:FFFF:FFFF



https://t.me/learningnets

For example, the solicited-node multicast address corresponding to the IPv6 address

2001:0DB8::1:800:200E:8C6C

is

FF02::1:FF0E:8C6C

IPv6 addresses that differ only in the high-order bits (for example, due to multiple high-order prefixes that are associated with different aggregations) will map to the same solicited-node address, which reduces the number of multicast addresses that a node must join.

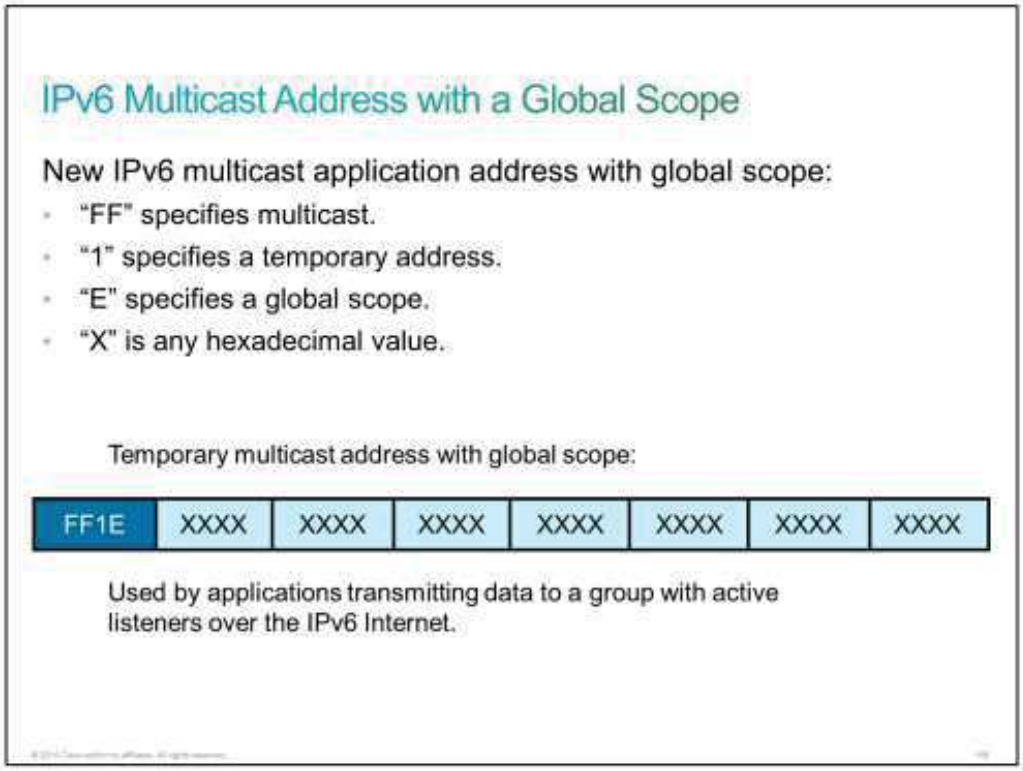
A node is required to compute and join (on the appropriate interface) the associated solicited-node multicast addresses for all unicast and anycast addresses that have been configured for the node interfaces (manually or automatically).



https://t.me/learningnets

# IPv6 Multicast Address with a Global Scope

The IPv6 multicast address with a global scope is described below:




When designing a new multicast application, it could be deployed on the Internet with a global scope at any address that meets this criterion:

```
FF1E:xxxx:xxxx:xxxx:xxxx:xxxx:xxxx:xxxx
```

The multicast address is composed of the following:

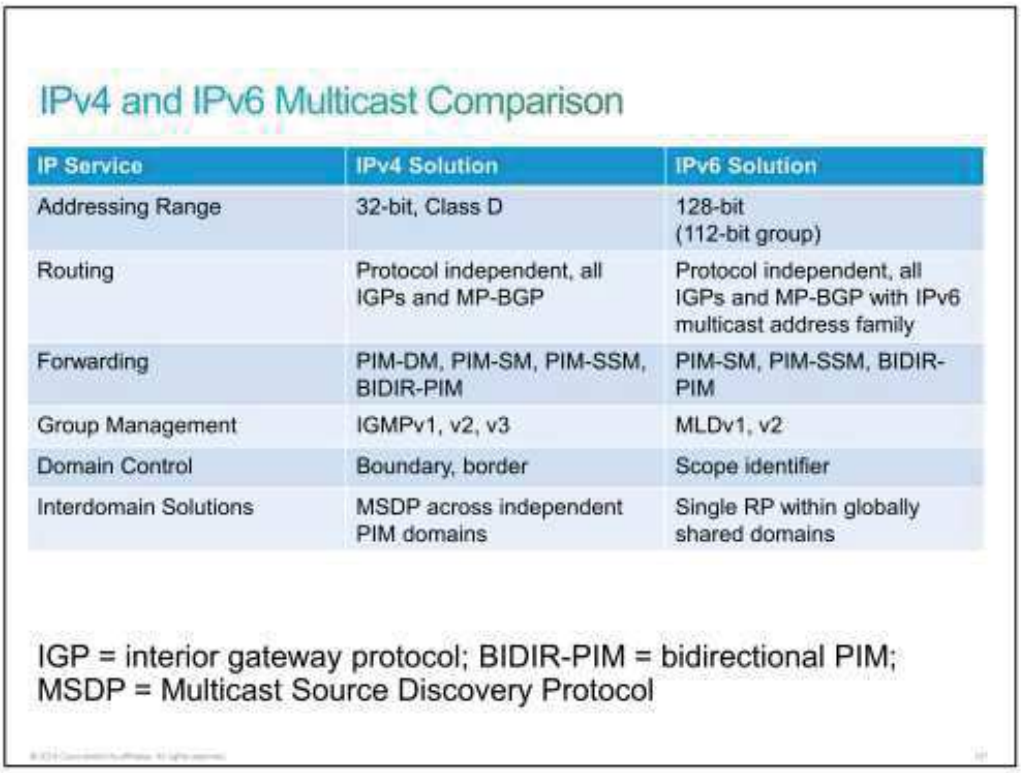
- "FF" specifies that this is a multicast address.
- "1" specifies a temporary address (not a reserved, permanent, IANA-registered multicast address).
- "E" specifies a global scope.
- "X" can be any hexadecimal value.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# IPv4 vs. IPv6 Multicast Comparison

This is IPv4 and IPv6 multicast comparison:




The most prominent changes with IPv6 multicast are the following:

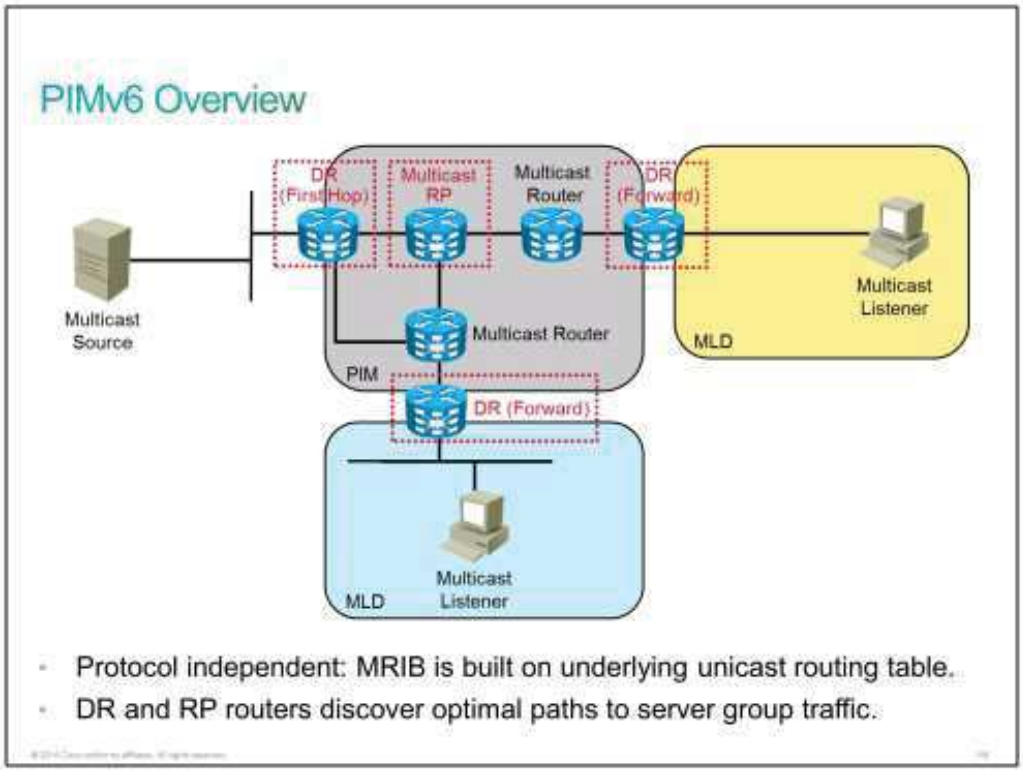
- IPv6 multicast address space is much larger.
- MP-BGP version 4 support allows population of separate routing paths for IPv6 unicast and multicast traffic.
- The Cisco IPv6 implementation does not support PIM-DM, but it is defined as a standard in *Protocol Independent Multicast–Dense Mode*, RFC 3973. However, due to resource consumption, dense mode is very rarely used.
- IPv6 multicast uses Multicast Listener Discovery (MLD), rather than Internet Group Management Protocol (IGMP) of IPv4, for group management (join and leave reports and queries).
- IPv6 multicast is scoped using IPv6 scoping rules.
- IPv6 uses a single RP for interdomain multicast.



https://t.me/learningnets

# PIMv6 Overview

PIMv6 is a routing protocol. PIMv6 is used between routers so that they can track which multicast packets to forward to each other and to their directly connected LANs. PIMv6 works independently of the unicast routing protocol to perform, send, or receive multicast route updates like other protocols. Regardless of which unicast routing protocols are used in the LAN to populate the unicast routing table, Cisco IOS PIMv6 uses the existing unicast table content to perform the RPF check instead of building and maintaining its own separate routing table.




You can configure IPv6 multicast to use either PIM-SM or PIM-SSM operations, or you can use both PIM-SM and PIM-SSM together in your network.

PIM relies on an underlying topology-gathering protocol to populate a routing table with routes. This routing table is called the MRIB. The routes in this table may be taken directly from the unicast routing table, or they may be different and provided by a separate routing protocol such as MBGP. Regardless of how it is created, the primary role of the MRIB in the PIM protocol is to provide the next-hop router along a multicast-capable path to each destination subnet. The MRIB is used to determine the next-hop neighbor to which any PIM join or prune message is sent. Data flows along the reverse path of the join message. Thus, in contrast to the unicast RIB, which specifies the next hop that a data packet would take to get to some subnet, the MRIB gives reverse-path information and indicates the path that a multicast data packet would take from its origin subnet to the router that has the MRIB.

MLD is used to allow link-local receivers to report interest in receiving specific multicast groups to their link-local router. PIM allows DRs serving multicast receivers (or last-hop routers), DRs serving multicast sources (or first-hop routers), and RPs to discover optimal forwarding paths for multicast control and data packets. The regions of the multicast network where PIM and MLD are active are shown in the figure. Note that routers can be first-hop routers for some multicast addresses and last-hop routers for other multicast addresses concurrently.



https://t.me/learningnets

"Protocol-independent" means that the multicast routing table can be built from the underlying unicast routing table. So regardless of whether the unicast routing table was built by RIPng, EIGRP, OSPFv3, or any other protocol, PIM can construct the multicast routing table. The only limitation in this case will be that multicast traffic and unicast traffic will follow the same path.

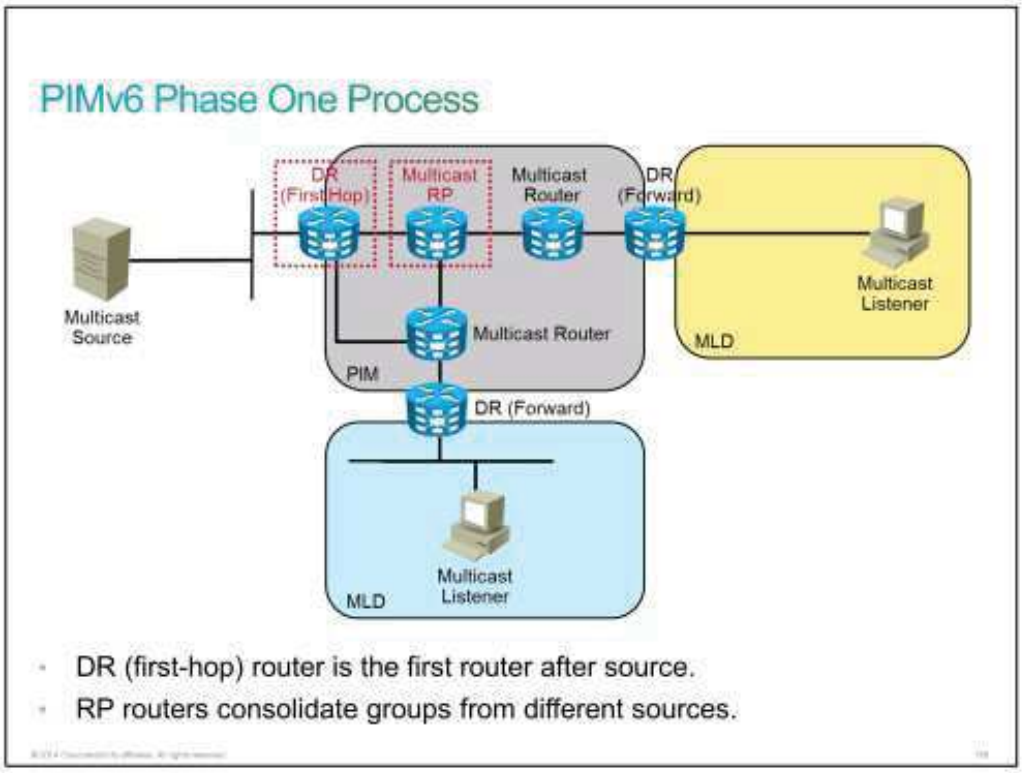
PIM also can run in addition to a multicast routing table that is built by a multicast-capable routing protocol like MP-BGP. In this case, the multicast traffic and unicast traffic can take different paths in the network, rather than those that are derived from the unicast routing table.

PIM packets are exchanged between all three of these main components:

- **DR (first hop) to and from RP:** This part of PIM enables new multicast sources to register with the RP that is associated with a given multicast address. This is done via PIM tunnels, in which the multicast packets are tunneled by the first-hop router to the RP. Later, the RP can obtain the multicast stream directly from the source. This is called building the SPT from the RP to the multicast source.
- **DR (last hop) to and from RP:** This part of PIM enables the DR router serving new multicast receivers to build a flooding path back to the RP through the intervening multicast routers in such a way that the DR can receive a specific stream. This results in the RP flooding packets toward the DR (last-hop) router that, in turn, results in a completed shared tree.
- **DR (last hop) to and from DR (first hop):** This is a process called building the SPT, in which the DR serving the listeners desires to get the multicast stream directly from the DR serving the multicast source (DR source).

Although the PIM process is a straightforward concept, PIM is a complex protocol.

## PIMv6 Phase One Process




Phase 1 of the network configuration for IPv6 multicast would be configuring the DR and RP routers.

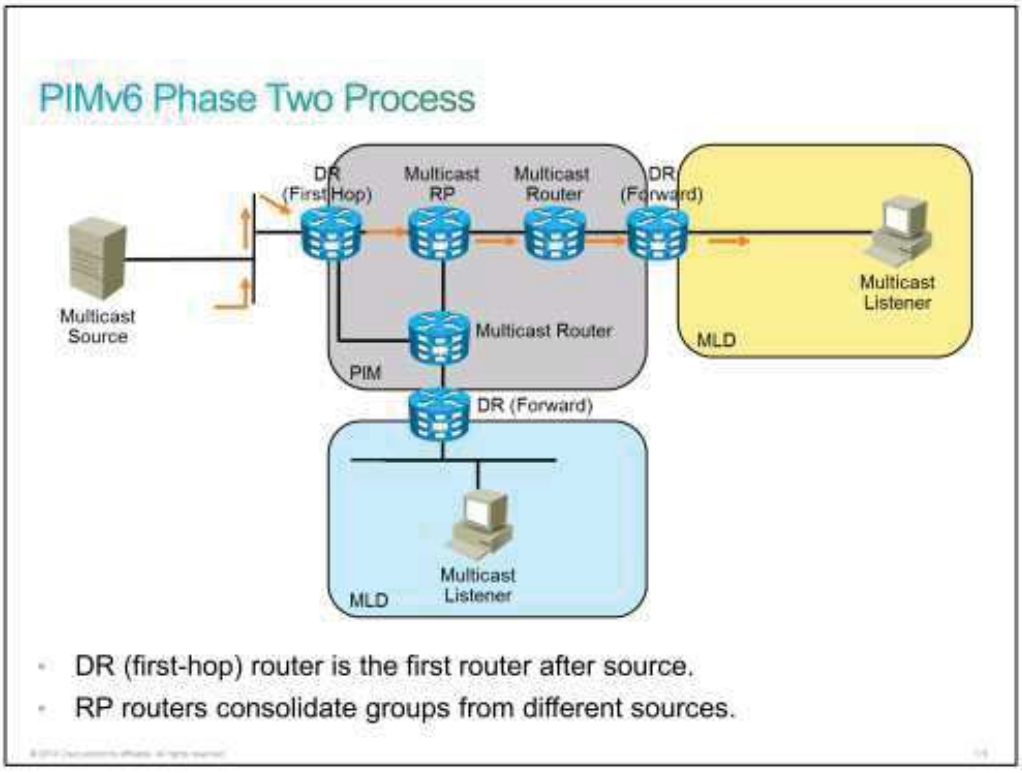


https://t.me/learningnets

Except for implementations that use techniques such as bootstrap router (BSR) or embedded RP, multicast routers are configured to know the location of the RP (such as IPv6 unicast or anycast address) for each multicast address group, and the RP is configured to know which multicast groups it will be serving as the RP for. This is static RP assignment.

In phase 1 of the PIM-SM protocol overview, a multicast receiver expresses its interest in receiving traffic that is destined for a multicast group. Typically, the multicast receiver does this using MLD. One of the local routers of the receiver is elected as the DR for this subnet. Upon receiving the expression of interest from the receiver, the DR then sends a PIM join message toward the RP for this multicast group. This join message is known as a (*,G) Join because it joins group G for all sources to this group. The (*,G) Join travels hop by hop toward the RP for the group, and in each router it passes through, multicast tree state for group G is instantiated. Eventually, the (*,G) Join either reaches the RP or reaches a router that already has the (*,G) Join state for this group. When many receivers join the group, their join messages converge on the RP and form a distribution tree for group G that is rooted at the RP. This is known as the RPT and is also known as the shared tree because it is shared by all sources sending to this group. Join messages are re-sent periodically as long as the receiver remains in the group. When all receivers on a leaf-network leave the group, the DR will send a PIM (*,G) Prune message toward the RP for this multicast group. However, if the prune message is not sent for any reason, the state will eventually time out.

## PIMv6 Phase Two Process




A multicast data sender just starts sending data that is destined for a multicast group. The sender local router (DR) takes those data packets, unicast encapsulates them, and sends them directly to the RP. The RP receives these encapsulated data packets, de-encapsulates them, and forwards them onto the shared tree. The packets then follow the (*,G) multicast tree state in the routers on the RP tree, being replicated wherever the RP tree branches, and eventually reaching all of the receivers for this multicast group. The process of encapsulating data packets to the RP is called registering, and the encapsulation packets are known as PIM register packets. At the end, multicast traffic is flowing encapsulated to the RP and then natively over the RPT to the multicast receivers.

Phase 2 of the PIM-SM protocol overview is register-stop.



https://t.me/learningnets

Register-encapsulation of data packets is inefficient for two reasons:

- Encapsulation and de-encapsulation may be relatively expensive operations for a router to perform, depending on whether the router has appropriate hardware for these tasks.
- Traveling all the way to the RP and then back down the shared tree may result in the packets traveling a relatively long distance to reach receivers that are close to the sender. For some applications, this increased latency or bandwidth consumption is undesirable.

Although register-encapsulation may continue indefinitely, for these reasons, the RP will normally choose to switch to native forwarding. To do this, when the RP receives a register-encapsulated data packet from source S on group G, it will normally initiate an (S,G) source-specific join toward S. This join message travels hop by hop toward S, instantiating the (S,G) multicast tree state in the routers along the path. The (S,G) multicast tree state is used only to forward packets for group G if those packets come from source S. Eventually, the join message reaches the S subnet or a router that already has the (S,G) multicast tree state, and then packets from S start to flow following the (S,G) tree state toward the RP. These data packets also may reach routers with the (*,G) state along the path toward the RP; if they do, they can shortcut onto the RPT at this point.

While the RP is in the process of joining the source-specific tree for S, the data packets will continue being encapsulated to the RP. When packets from S also start to arrive natively at the RP, the RP will receive two copies of each of these packets. At this point, the RP starts to discard the encapsulated copy of these packets, and it sends a register-stop message back to the S DR to prevent the DR from unnecessarily encapsulating the packets.

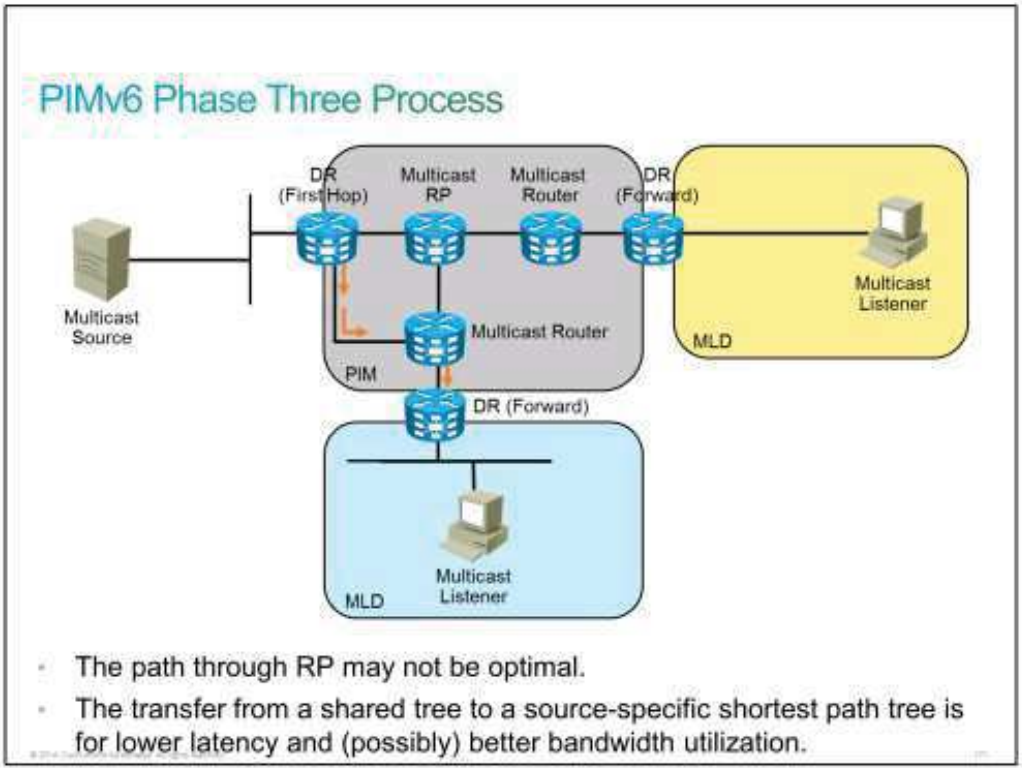
At the end of phase 2, traffic will flow natively from S along a source-specific tree to the RP and from there along the shared tree to the receivers. Where the two trees intersect, traffic may transfer from the source-specific tree to the RPT and thus avoid taking a long detour via the RP.

| **Note** | A sender may start sending before or after a receiver joins the group, and thus phase 2 may happen before the shared tree to the receiver is built. |
|---|---|



https://t.me/learningnets

# PIMv6 Phase Three Process




Although having the RP join back toward the source removes the encapsulation overhead, it does not completely optimize the forwarding paths. For many receivers, the route via the RP may involve a significant detour when compared with the shortest path from the source to the receiver.

To obtain lower latencies or more efficient bandwidth utilization, a router on the receiver LAN, typically the DR, may optionally initiate a transfer from the shared tree to a source-specific SPT. To do this, it issues an (S,G) Join toward S. This instantiates state in the routers along the path to S. Eventually, this join either reaches the S subnet, or it reaches a router that already has the (S,G) state. When this happens, data packets from S start to flow following the (S,G) state until they reach the receiver.

PIM is used to manage the flow of these packets, pruning the stream back when the last listener has stopped listening to a specific multicast group.

At this point, the receiver (or a router upstream of the receiver) will receive two copies of the data: one from the SPT and one from the RPT. When the first traffic starts to arrive from the SPT, the DR or upstream router starts to drop the packets for G from S that arrive via the RPT. In addition, it sends an (S,G) prune message toward the RP. This is known as an (S,G,rpt) Prune. The prune message travels hop by hop, instantiating state along the path toward the RP, indicating that traffic from S for G should not be forwarded in this direction. The prune is propagated until it reaches the RP or a router that still needs the traffic from S for other receivers.

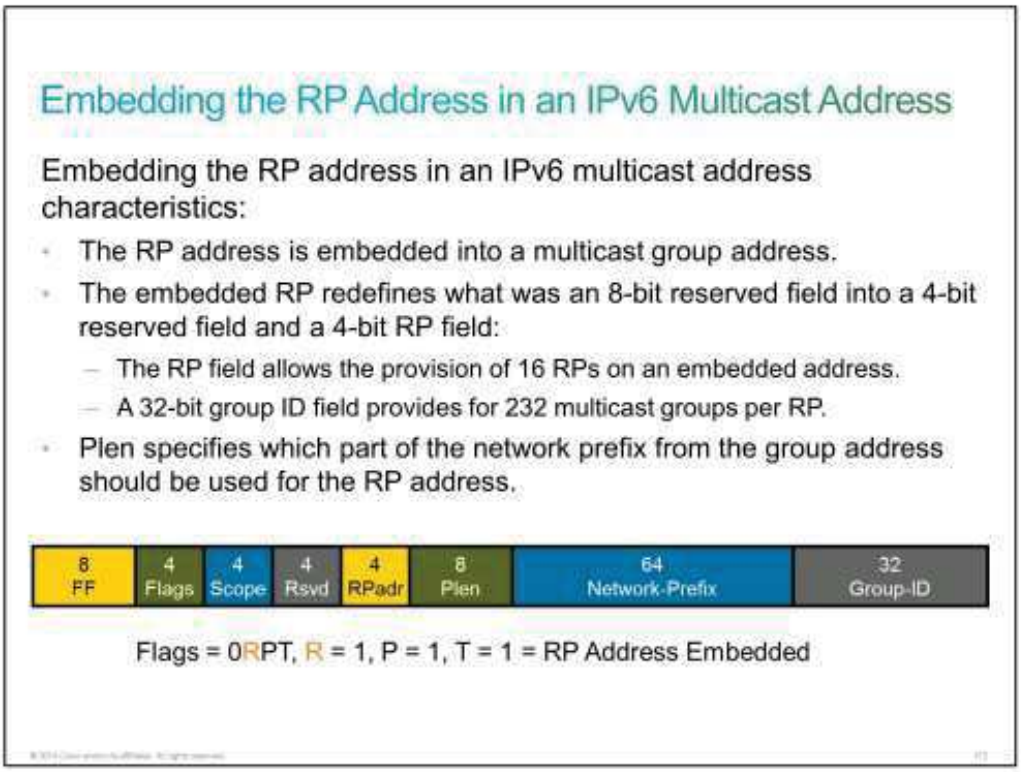
By now, the receiver will receive traffic from S along the SPT between the receiver and S. In addition, the RP receives the traffic from S, but this traffic is no longer reaching the receiver along the RPT. As far as the receiver is concerned, this is the final distribution tree.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# Embedding the RP Address in an IPv6 Multicast Address

In static RP, all PIM-enabled routers are statically configured with the RP IPv6 address that should be used for a given multicast group. Whenever the forwarding router discovers a new listener on an attached link for a new multicast group, the router builds state toward this RP. This implies that the first-hop DR and all other multicast-enabled routers in the domain are configured to register with this specific RP for multicast streams.




Embedded RP is a mechanism that eliminates the need for static configuration of RPs on the multicast DRs. The RP address for a given multicast group (address) is encoded inside the multicast group.

The figure shows the format for the multicast group address. It is derived via the mechanism that is described in RFC 3956, *Embedding the Rendezvous Point (RP) Address in an IPv6 Multicast Address*, and is also a unicast-prefix multicast group address. A multicast group address uses these special rules, with fields described from left to right:

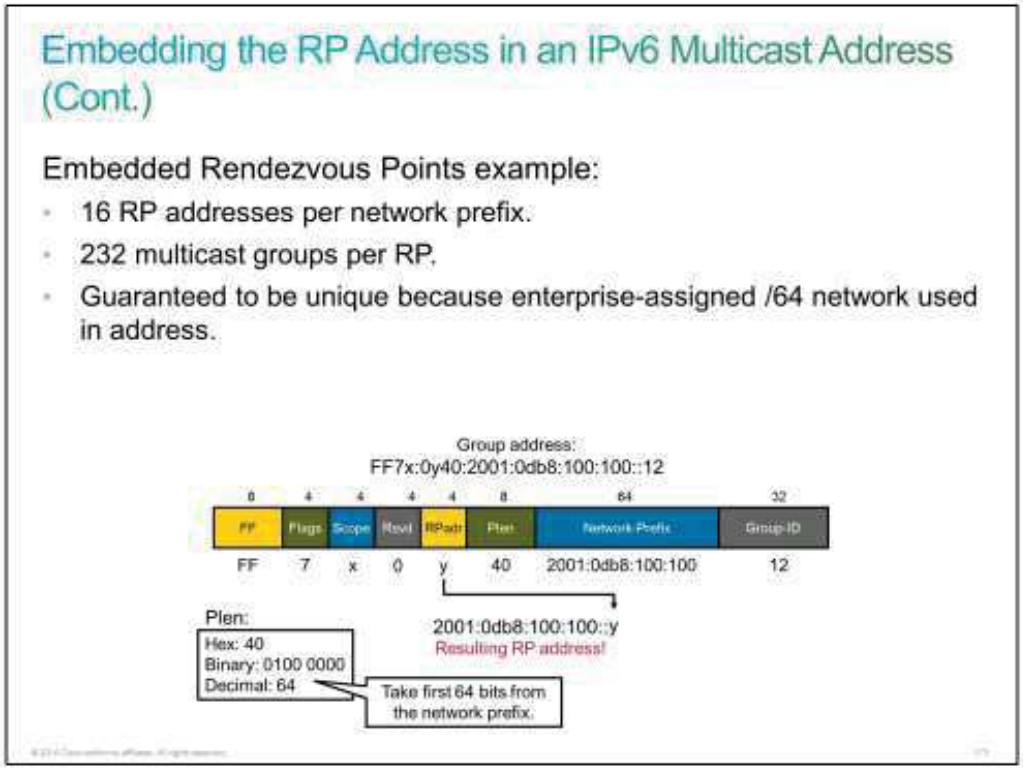
- **FF:** This is a multicast designator.
- **Marks:** This field contains 4 bits, which is defined from left to right as follows:
  - **First bit:** This is reserved.
  - **Second bit:** This is known as the R field, which, when set to 1, specifies that this multicast address contains an embedded RP address.
  - **Third bit:** This is known as the P field, which specifies if the address contains a unicast prefix-based address. It is set to 1 if it does contain a unicast prefix-based address. The unicast prefix included in the address is for the multicast source home prefix, making this multicast group globally unique. If the R bit is set to 1, the P bit must be set to 1 as well.



https://t.me/learningnets

- **Fourth bit:** This is known as the T field, and it is set to 0 if this is a permanently-assigned, well-known multicast address and to 1 if temporary. For all cases of embedded RP, the T bit must be set to 1.

- **Scope:** This field has the same values as for any multicast address.
- **Rsvd (reserved):** This field is commonly set to all zeros. If this was not an embedded RP multicast address, this would be an 8-bit reserved field.
- **RPadr:** This is a 4-bit field that specifies RPs 1 to 16 on the /64 network prefix embedded farther along in the address.
- **Plen (prefix length):** This value must not be 0 (SSM) or greater than 64. The value specifies how many bits of network prefix from the group address will be used to determine the RP address.
- **Network-Prefix:** This field is the network prefix for the RP.
- **Group-ID:** This is the multicast group ID, which is either a well-known ID, such as NTP, or a temporary ID that is invented by an enterprise. Remember that the marks field identifies the address as well known or transient.




This example shows the resulting RP address that is taken from the multicast group address. The 4 bits taken from the previously reserved range are used to specify one of 16 addresses, such as 1–15 (0 is reserved). Therefore, an address ending in ::16 or larger could not be an embedded RP.

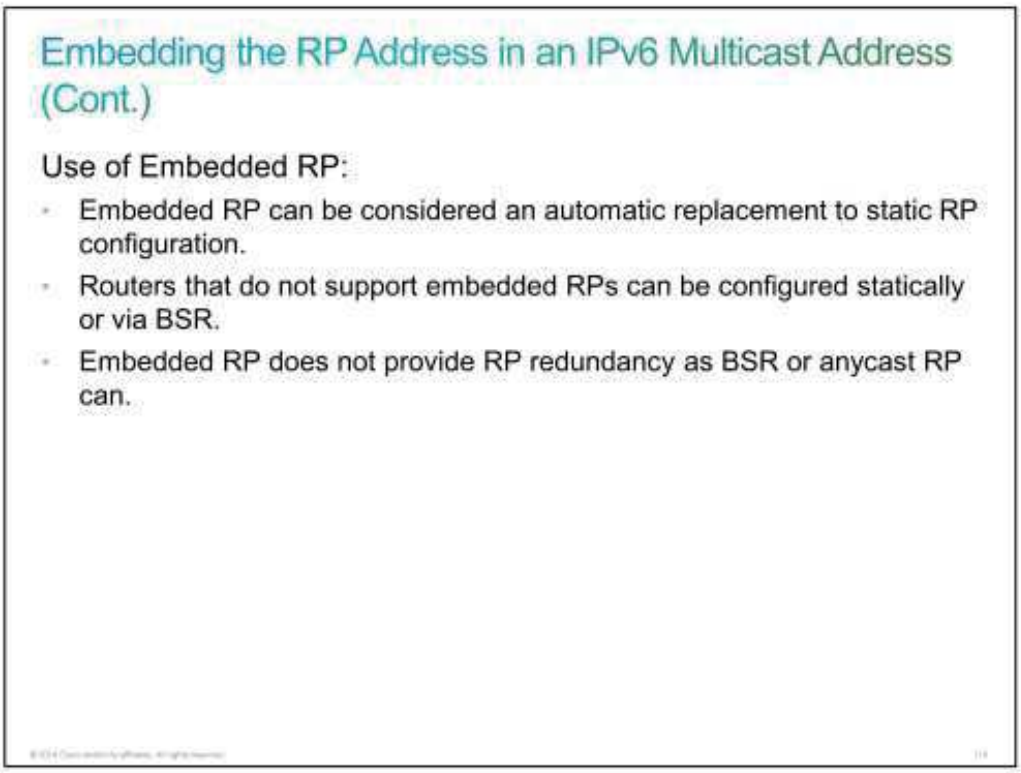
In the example, the mark bits are set to "0111," which means "embedded RP, unicast-based, temporary" multicast group. Plen is set to 40, which means that 64 (all) bits of the network prefix are used to determine the RP IP address. Group ID in the example is set to 12, and the RP address ("y") can range from 1 to 15.

Using prefix embedding, any enterprise with a /64-bit prefix can generate a unique multicast address (or group)—guaranteed to be unique Internet-wide—without consulting any organization or registering with IANA.



https://t.me/learningnets

Embedded RP adds the ability to also include 16 possible addresses for RPs for the multicast group in the group ID (or multicast address). Because the RP address is encoded in the address and located within the IPv6 network of the "inventing" enterprise, this enterprise is able to set up an RP to service its new multicast group. The enterprise can set up a maximum of 16 RPs for this multicast group.




Embedded RP is ideal as a replacement for static RP. No manual configuration of RP addresses is required because the address is embedded within a multicast group. However, if the RP address needs to change, the multicast address needs to change, which requires informing all interested applications. Routers that do not support embedded RP can be statically configured or use other methods such as BSR.

| | |
|---|---|
| **Note** | The Cisco Auto-RP does not support IPv6. |

Routers that serve as embedded RPs must be configured for this specific multicast group. This process provides a safeguard mechanism that prohibits poor selection of RPs. By specifying exactly which multicast groups that an RP will stream traffic for, there is less likelihood that a high-rate group address will be sourced from a low-throughput RP. Without this safeguard, you could select RPs simply by specifying the multicast address with an embedded RP address, without consulting the network team, resulting in a possibility that an embedded RP address could specify an RP that could not process the multicast traffic (for example, an RP that is designed only for low-throughput multicast).

| | |
|---|---|
| **Note** | The embedded RP solves the problem of helping multicast routers choose the right RP, but it does not address the redundancy of RPs if the RP goes down. |

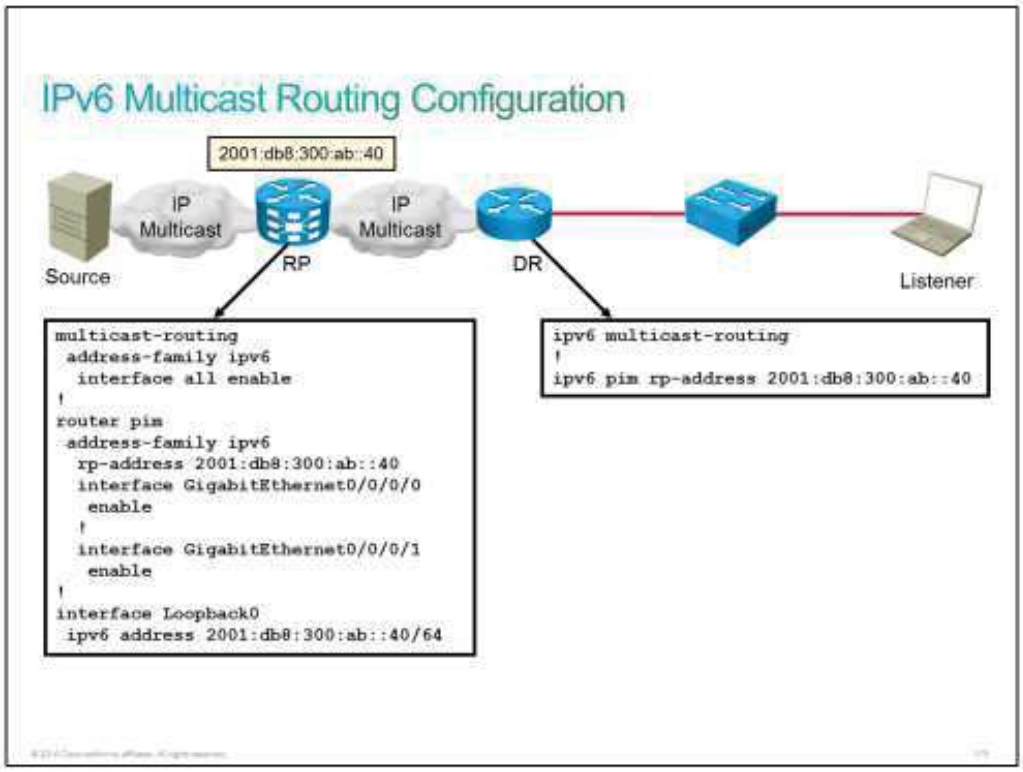
Embedded RP is an IPv6-only multicast feature, which simplifies the process of managing mapping between multicast groups and RPs.



https://t.me/learningnets

# IPv6 Multicast Routing Configuration

The figure shows basic multicast configuration for IPv6 on Cisco IOS XR, IOS, and IOS XE Software platforms.




On the Cisco IOS XR Software platform, first enter the multicast routing configuration mode using the **multicast-routing** command. This command starts the following multicast processes: MRIB, MFWD, muRIB, PIM-SM, IGMP, and MLD. Then enter the address family configuration mode for IPv6. You can enable multicast on relevant interfaces using the **interface** *interface* **enable** command. You can also enable multicast on all interfaces using the **interface all enable** command. To statically configure RP, enter router PIM configuration mode and specify RP using the **rp-address** command. Then enter interface configuration mode and enable PIM on an interface using the **enable** command. In the example, the Cisco IOS-XR router will act as an RP; therefore, a loopback interface is also configured. The loopback is configured with an IPv6 address that will be used as the RP.

On the Cisco IOS and IOS XE Software platforms, enable multicast for IPv6 using the **ipv6 multicast-routing** command. Then statically specify RP using the **ipv6 pim rp-address** command.

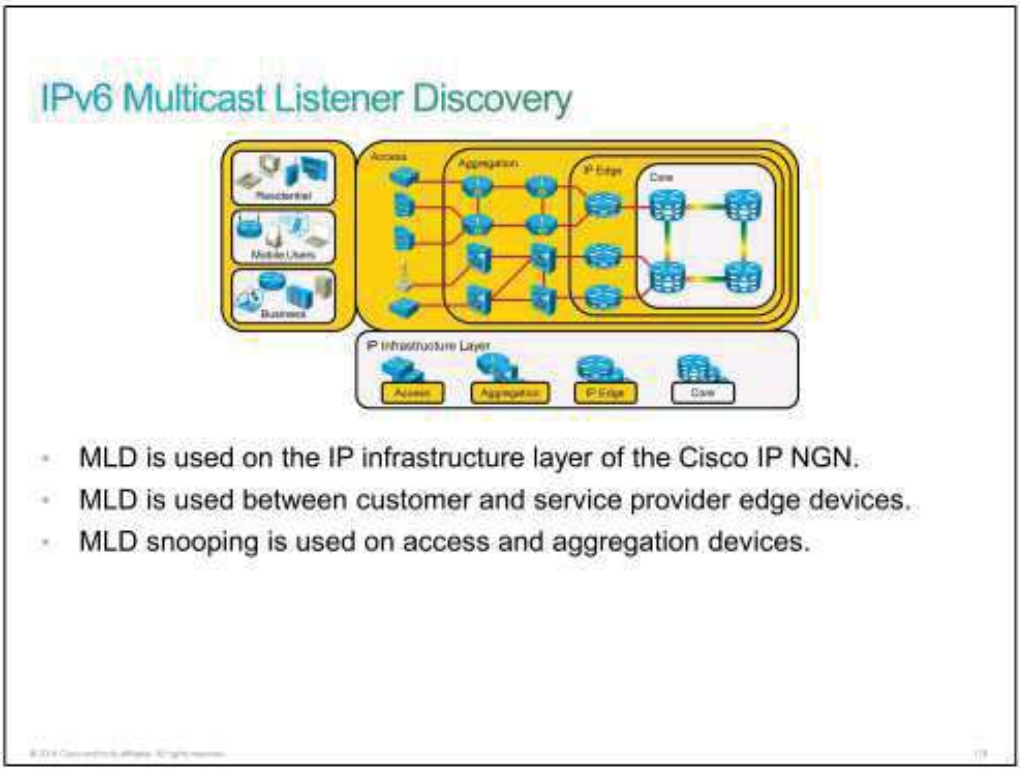
| | |
|---|---|
| **Note** | For a complete command reference, refer to the Cisco IOS, IOS XE, and IOS XR Software Command Reference guides on www.cisco.com. |



https://t.me/learningnets

# IPv6 Multicast Listener Discovery

MLD is used in the infrastructure layer of the Cisco IP NGN. Multicast is used between customer edge CE and PE devices.




MLD snooping is used on access and aggregation devices as well.

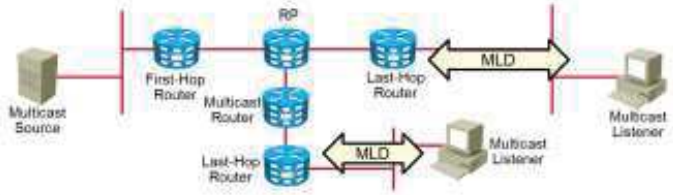


https://t.me/learningnets

## IPv6 Multicast Listener Discovery (Cont.)

IPv6 MLD characteristics:

- To implement multicast, routers should know which users should receive the multicast traffic.
- IPv4: IGMP
- IPv6: MLD
  - MLDv1 is similar to IGMPv2
  - MLDv2 is similar to IGMPv3
- MLD handles join and leave processes on the access segment between the listener and the first multicast router.




The MLD protocol is used on the LAN segment that interconnects the listeners and a multicast router. Using MLD, the listeners and the router control the participation of the listeners in multicast groups.

To start implementing multicasting in the campus network, users must first define who receives the multicast.

MLD specifies the protocol that is used by an IPv6 multicast router to discover the presence of multicast receivers (or nodes wishing to receive multicast packets) on its directly attached links and to discover specifically which multicast addresses are of interest to those neighboring nodes. This information is then provided to whichever multicast routing protocol (for example, PIM) is used by the router to ensure that multicast packets are delivered to all links in which there are interested receivers. The MLD protocol provides a means to automatically control and limit the flow of multicast traffic throughout the network with the use of special multicast queriers and hosts.

The difference between multicast queriers and hosts is as follows:

- A querier is a network device, such as a router, that sends query messages to discover which network devices are members of a given multicast group.
- A host is a receiver, including a router, that sends report messages to inform the querier of a host membership.

A set of queriers and hosts that receive multicast data streams from the same source is called a multicast group. Queriers and hosts use MLD reports to join and leave multicast groups and to begin receiving group traffic.

MLD is an asymmetric protocol, specifying different behaviors for multicast receivers and routers. A node, for example, always performs the "listener" part of MLD. This means that the node wants to listen to specific multicast streams and needs to notify the upstream router about what those specific addresses are. The router can perform the listener part of MLD as well (and it also can want to listen to specific multicast addresses), but the router always performs the "router" part of MLD.



https://t.me/learningnets

MLDv1 is equivalent to IPv4 multicast IGMPv2, where MLD messages are always between multicast routers and link-local multicast nodes. So MLD messages always use link-local addressing. One important difference to note is that MLD uses ICMPv6 (IP Protocol 58) message types rather than IGMP (IP Protocol 2) message types.

MLD messaging uses the IPv6 router alert header option (within a hop-by-hop option, as described in RFC 2711, IPv6 Router Alert Option) to signal to the MLD router to look into these packets because some hop-by-hop service is needed. Use of the router alert option is needed to force the MLD router to look into packets that are not destined to itself.

There are two versions of MLD today: MLDv1 (which is based on IGMPv2) and MLDv2 (which is based on IGMPv3). MLDv2 provides support for link-local receivers to specify sources for multicast groups—called source filtering—and to specify just the multicast group. This, in turn, provides support for SSM, where the last-hop router can obtain the multicast stream directly from the multicast source first-hop router.

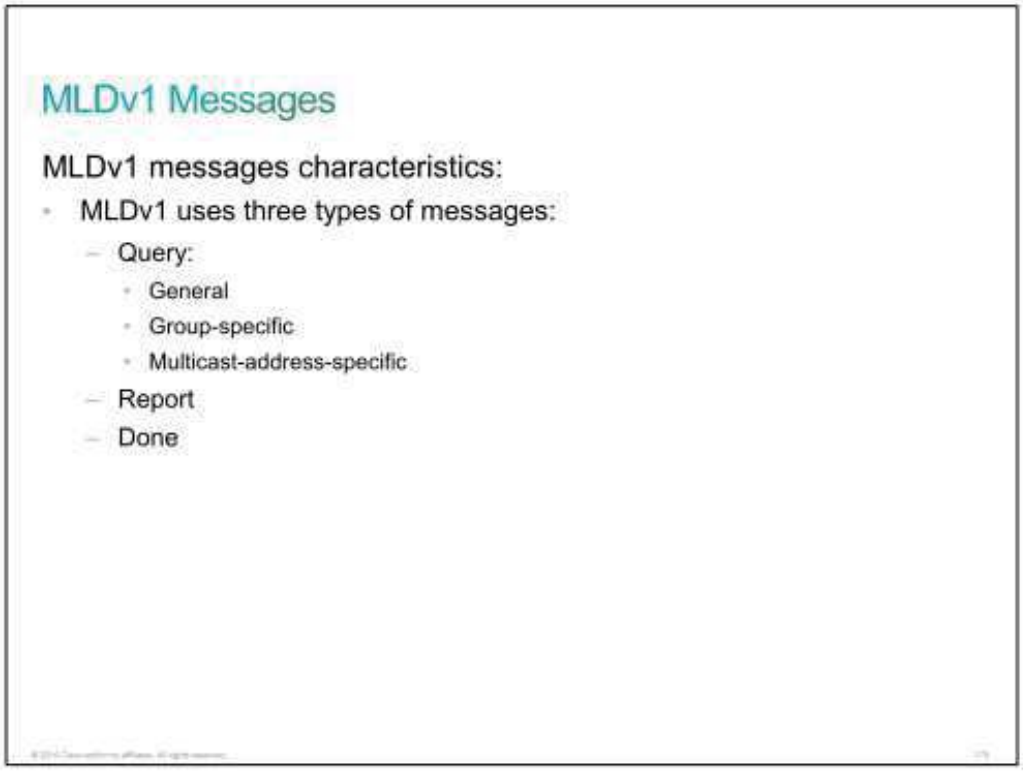
In an enterprise scenario, switching mechanisms causes multicast packets to be sent to ports that are not multicast listeners because the switch has no way to determine which multicast listeners are on which ports. MLD snooping, described in RFC 4541, *Considerations for IGMP and MLD Snooping Switches*, describes how switches can listen to MLD packets to determine how to more efficiently forward multicast packets. This is done in hardware by high-end switches in most cases.



https://t.me/learningnets

# MLDv1 Messages

The MLDv1 messages are described below.




MLDv1 defines three types of messages:

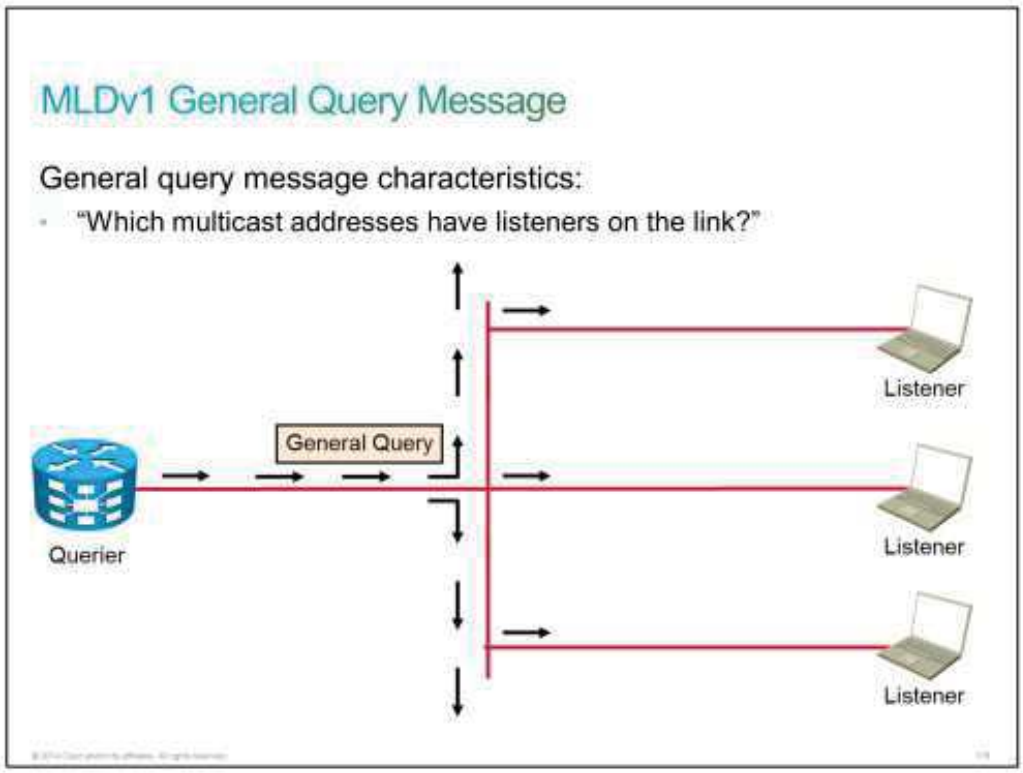
- **Query:** In a general, group-specific, and multicast-address-specific query message, the multicast address field is set to 0 when MLD sends a general query. The general query learns which multicast addresses have listeners on an attached link. Group-specific and multicast-address-specific queries are the same. A group address is a multicast address.
- **Report:** In a report message, the multicast address field is that of the specific IPv6 multicast address to which the sender is listening.
- **Done:** In a done message, the multicast address field is that of the specific IPv6 multicast address to which the source of the MLD message is no longer listening.



https://t.me/learningnets

# MLDv1 General Query Message

Routers use MLD to learn which multicast addresses have listeners on each of their attached links. Each router keeps a list for each attached link, of which multicast addresses have listeners on this link and a timer that is associated with each of those addresses. Note that the router needs to learn only that listeners for a given multicast address are present on a link. The router does not need to learn the identity (for example, unicast address) of those listeners or even how many listeners are present.




MLDv1 routers send general queries on each link, asking for reports from each multicast link-local receiver (node). This information tells the DR for the link which multicast groups the link-local receivers are listening to.

For each attached link, a router selects one of its link-local unicast addresses on this link to be used as the IPv6 source address in all MLD packets that the router transmits on this link.

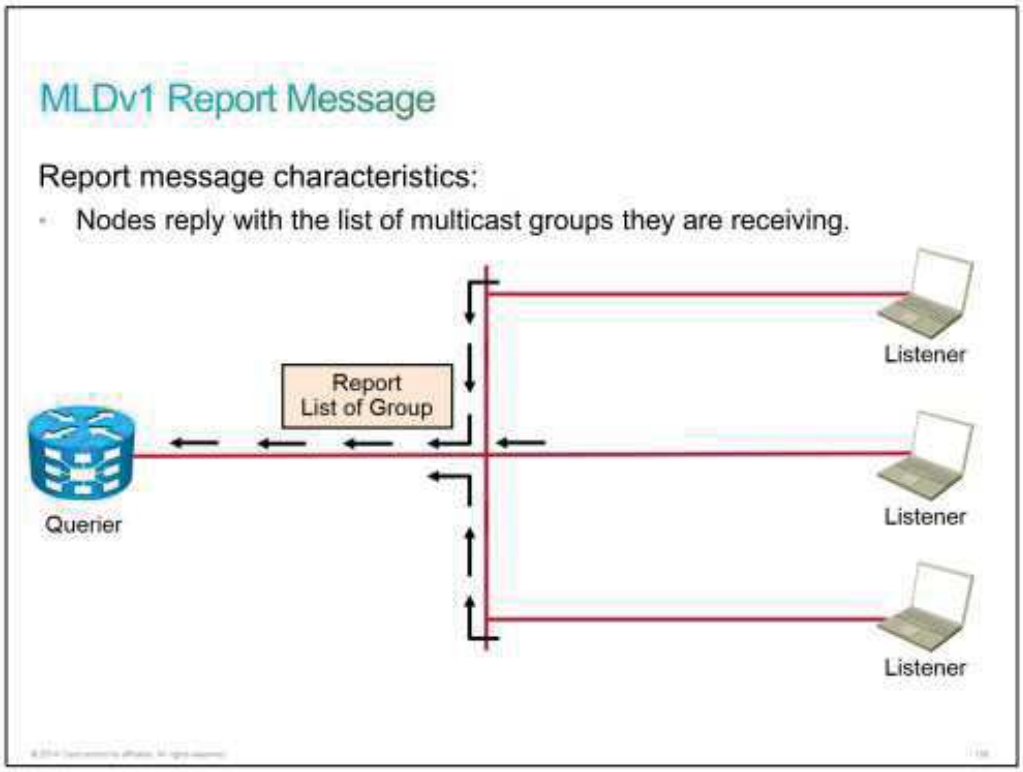
General queries are sent to the link-scope all-nodes multicast address (FF02::1), with a Multicast Address field of 0 and a Maximum Response Delay of Query Response Interval.



https://t.me/learningnets

# MLDv1 Report Message

When a node receives a general query, it sets a delay timer for each multicast address to which it is listening on the interface from which it received the query, excluding the link-scope, all-nodes address and any multicast addresses of scope 0 (reserved) or 1 (node-local). Each timer is set to a different random value, using the highest clock granularity that is available on the node, selected from the range (0, Maximum Response Delay) with Maximum Response Delay as specified in the query packet. If a timer for any address is already running, it is reset to the new random value only if the requested Maximum Response Delay is less than the remaining value of the running timer. If the query packet specifies a Maximum Response Delay of 0, then each timer is effectively set to 0, and the action that is specified for timer expiration is performed immediately.




If a node timer for a particular multicast address on a particular interface expires, the node transmits a report to this address via this interface; the address being reported is carried in both the IPv6 Destination Address field and the MLD Multicast Address field of the report packet. The IPv6 hop limit of 1 (and the presence of a link-local IPv6 source address) prevent the packet from traveling beyond the link to which the reporting interface is attached.

If a node receives another node report from an interface for a multicast address while the node has a timer running for this same address on this interface, it stops its timer and does not send a report for this address, thus suppressing duplicate reports on the link.

When a node starts listening to a multicast address on an interface, it should immediately transmit an unsolicited report for this address on this interface if it is the first listener on the link. To cover the possibility of the initial report being lost or damaged, it is recommended that it be repeated once or twice after short delays (that is, Unsolicited Report Interval). A simple way to accomplish this is to send the initial report and then act as if a multicast-address-specific query was received for this address, and set a timer appropriately.

These are unsolicited reports.

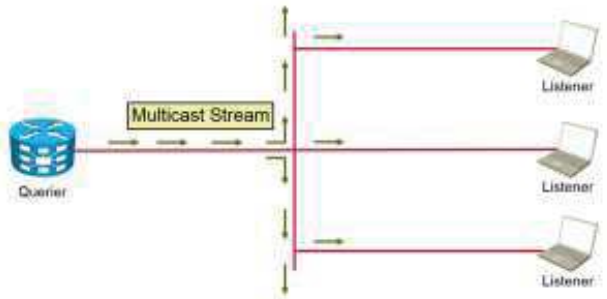


https://t.me/learningnets

## MLDv1 Report Message (Cont.)

General report MLDv1 process:

- In response to a MLD report, the router procures itself the traffic for a multicast group (for example, using a PIM join message).
- The router starts transmitting the group data to the listener on the link.




When a router receives a report from a link, if the reported address is not already present in the router list of a multicast address having listeners on this link, the reported address is added to the list, its timer is set to Multicast Listener Interval, and its appearance is made known to the router multicast routing component. If a report is received for a multicast address that is already present in the router list, the timer for this address is reset to Multicast Listener Interval. If an address timer expires, it is assumed that there are no longer any listeners for this address that is present on the link, so this address is deleted from the list, and its disappearance is made known to the multicast routing component.

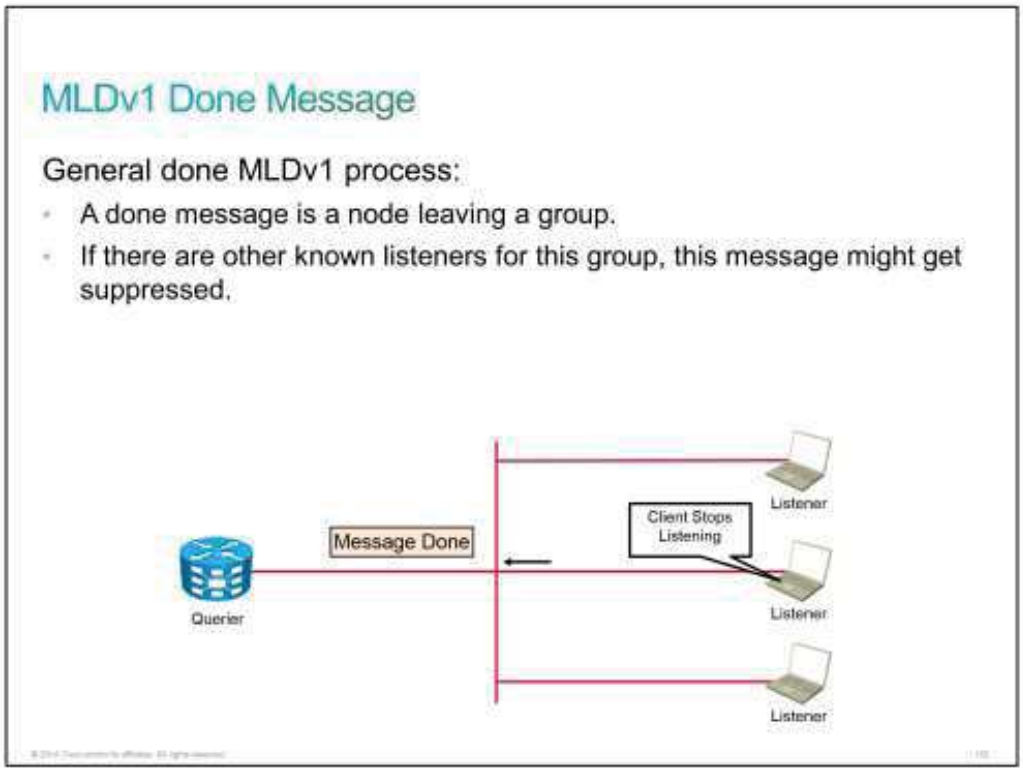
Each multicast stream is obtained independently. When the stream arrives, the router floods the packets to the appropriate link.



https://t.me/learningnets

# MLDv1 Done Message

When a node ceases to listen to a multicast address on an interface, it should send a single done message to the link-scope, all-routers multicast address (FF02::2), carrying in its multicast address field the address to which it is ceasing to listen.



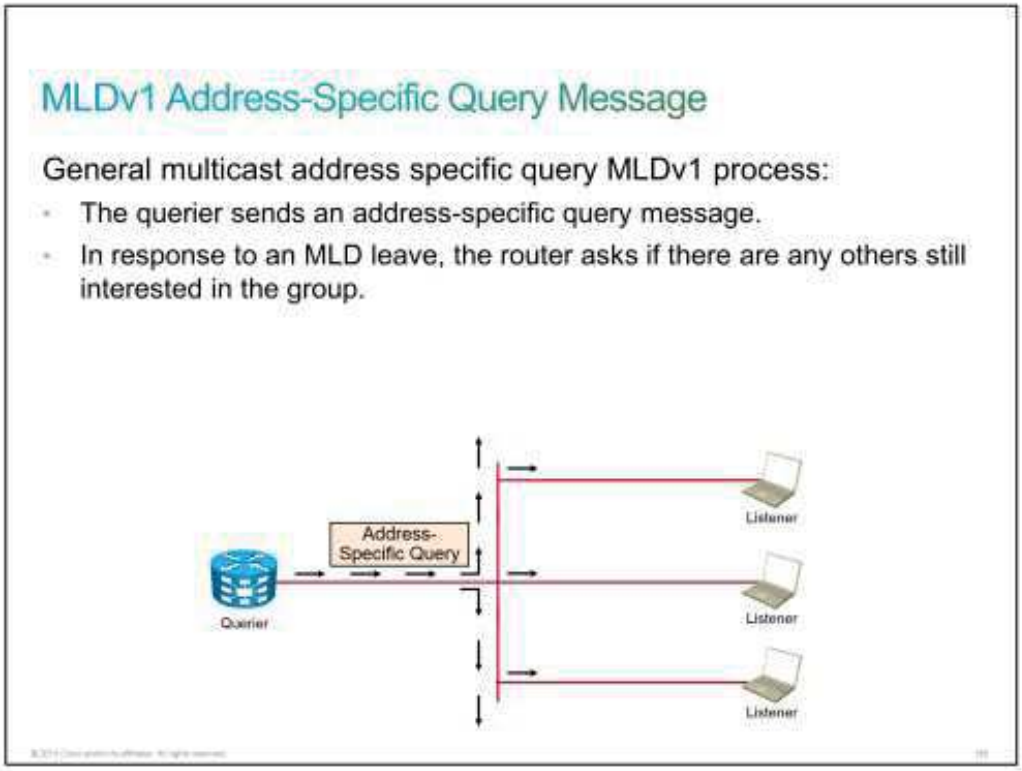

If the most recent report message of the node was suppressed by hearing another report message, it may send nothing, as it is highly likely that there is another listener for this address that is still present on the same link.



https://t.me/learningnets

# MLDv1 Address-Specific Query Message

When a router in querier state receives a done message from a link, if the multicast address that is identified in the message is present in the querier list of addresses having listeners on this link, the querier sends multiple multicast-address-specific queries (based on the Last Listener Query Count) once every interval (based on the Last Listener Query Interval) to this multicast address.



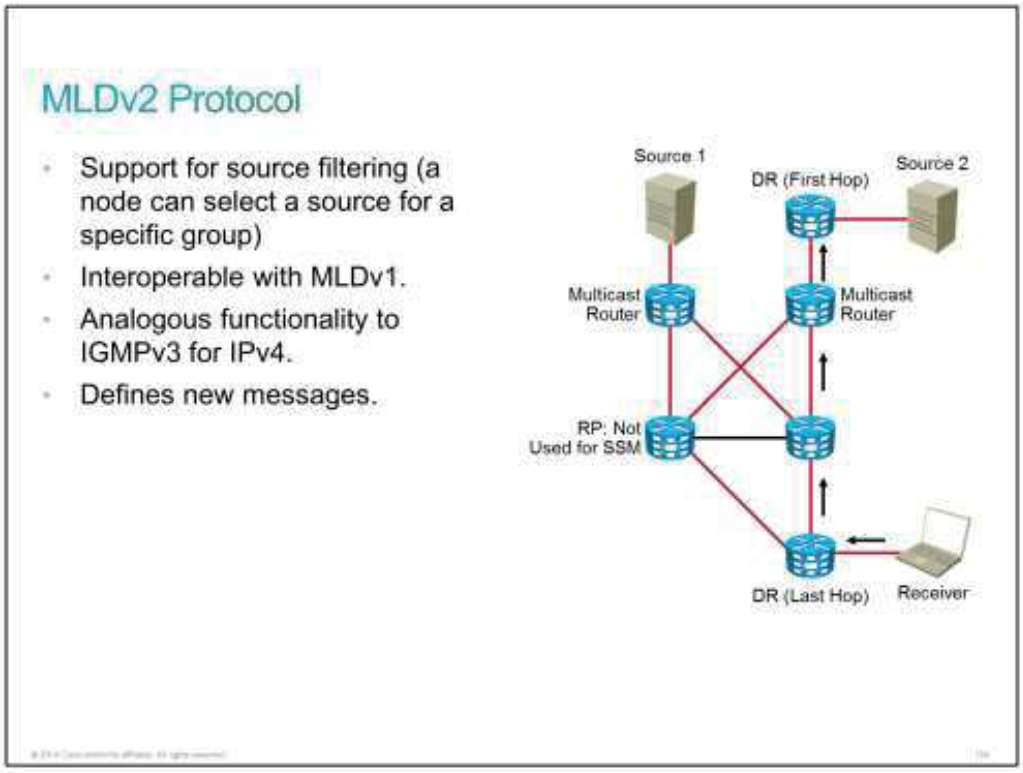

These multicast-address-specific queries have their Maximum Response Delay set to Last Listener Query Interval. If no reports for the address are received from the link after the response delay of the last query has passed, then the routers on the link assume that the address no longer has any listeners there; the address is therefore deleted from the list, and its disappearance is made known to the multicast routing component. This process is continued to its resolution (that is, until a report is received or the last multicast-address-specific query is sent with no response) despite any transition from querier to nonquerier on this link.



https://t.me/learningnets

# MLDv2 Protocol

MLD is used by an IPv6 router to discover the presence of multicast listeners on directly attached links, and MLD is used to discover which multicast addresses are of interest to those neighboring nodes. MLDv2 is interoperable with MLDv1. MLDv2 adds the ability for a node to report interest in listening to packets with a particular multicast address only from specific source addresses or from all sources except for specific source addresses.




MLDv2 is a translation of the IGMPv3 protocol (RFC 3376) for IPv6 semantics.

The MLDv2 protocol, when compared to MLDv1, adds support for "source filtering," which means a node is able to report interest in listening to packets from specific source addresses, as required to support source-specific multicast (RFC 3569), or from specific source addresses that are sent to a particular multicast address. MLDv2 is interoperable with MLDv1.

A node maintains or computes multicast listening state for each of its interfaces. Conceptually, this state consists of a set of records, with each record containing an IPv6 multicast address, a filter mode, and a source list. The filter mode may be either INCLUDE or EXCLUDE. In INCLUDE mode, reception of packets that are sent to the specified multicast address is enabled only from the source addresses that are listed in the source list. In EXCLUDE mode, reception of packets that are sent to the given multicast address is enabled from all source addresses except those listed in the source list.

For example, with MLDv1, a router has to track multicast destinations for which there are link-local listeners. A multi-interface router needs to track this for each interface and build a consolidated set of destination addresses that it wishes to be sent to the router. For MLDv2, a router must track, for each link, the list of multicast destination addresses and the list of associated source addresses for those multicast destinations, and a router must keep a consolidated list of those streams. This additional task makes it possible to fully specify, as a multicast receiver, a very specific multicast stream to listen for. Even if two multicast senders are using the same multicast destination address (not a reserved but a self-selected temporary address), a receiver could be sure to listen to the desired stream.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

In the example above, the receiver indicates an interest in a given multicast group, via MLDv2, to the DR. The MLDv2 report includes, for SSM, the source addresses for the group. The router uses PIM to build forwarding state directly back to the source; no RP is involved. In this way, MLDv2 enables SSM. If no source is specified, multicast state is built back to the RP as in any-source multicasting.

There are three types of MLDv2 query messages:

- General queries
- Multicast-address-specific queries
- Multicast-address and source-specific queries

The querier periodically sends general queries to learn multicast address listener information from an attached link. These queries are used to build and refresh the multicast address listener state inside all multicast routers on the link.

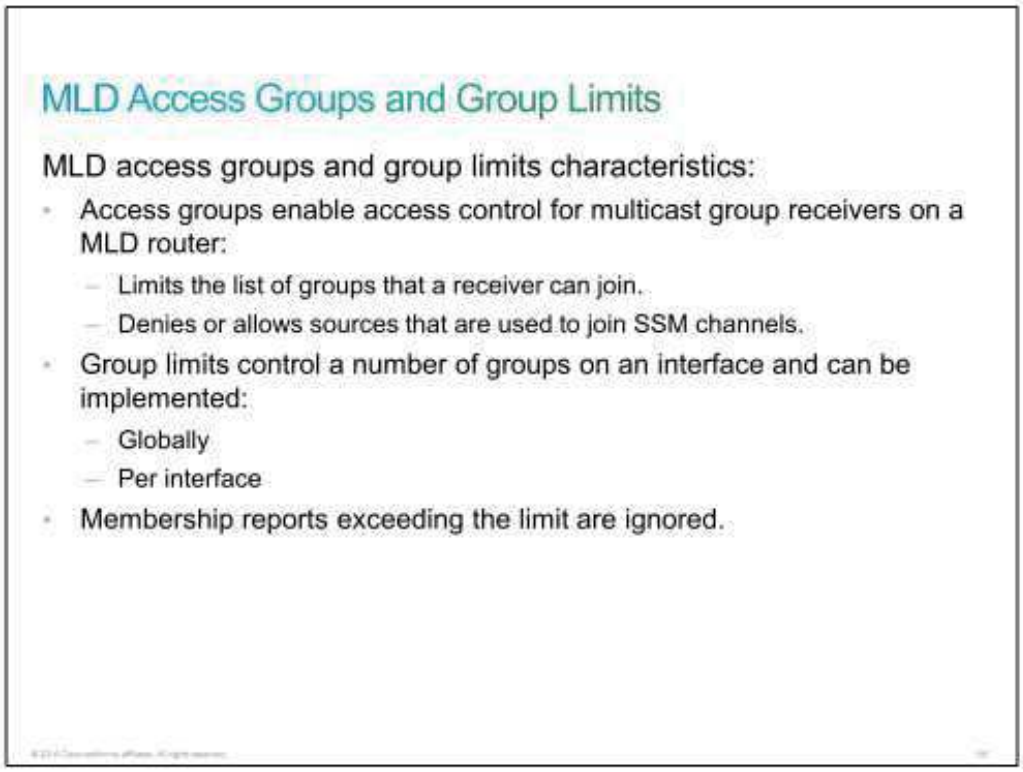
Nodes respond to these queries by reporting their per-interface multicast address listener state through current state report messages that are sent to a specific multicast address that all MLDv2 routers on the link listen to. On the other hand, if the listening state of a node changes, the node immediately reports these changes through a state change report message. The state change report contains filter mode change records, source list change records, or records of both types.



https://t.me/learningnets

# MLD Access Groups and Group Limits

MLD access groups provide receiver access control in IPv6 multicast routers. This feature limits the list of groups that a receiver can join, and it allows or denies sources that are used to join SSM channels.



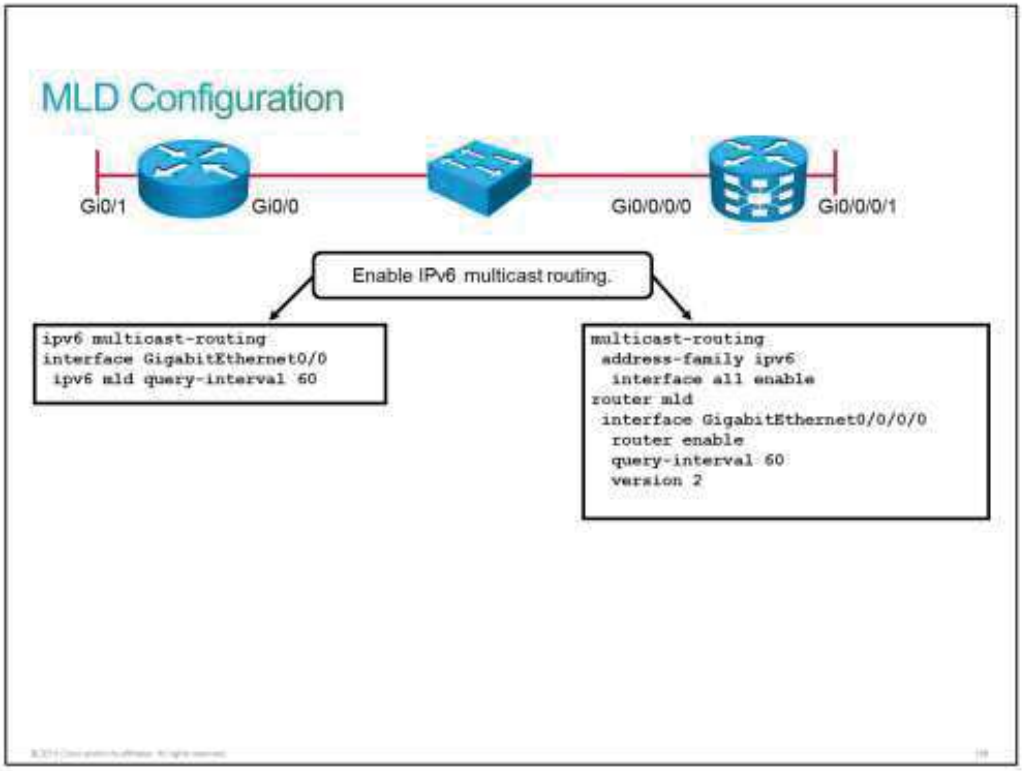

MLD group limits control a number of groups on an interface and can be implemented globally or per interface. Per-interface and global MLD limits operate independently of each other. Both per-interface and global MLD limits can be configured on the same router. The number of MLD limits, globally or per interface, is not configured by default; the limits must be configured by the user. A membership report that exceeds either the per-interface or the global state limit is ignored.



https://t.me/learningnets

# MLD Configuration

On the Cisco IOS and IOS XE Software platforms, MLD is enabled by default when IPv6 multicast routing is enabled (with the **ipv6 multicast-routing** command). On the Cisco IOS XR Software router, you have to enable IPv6 multicast routing and enable **address-family ipv6** for the interface. Use the following Cisco IOS XR Software commands to enable multicast routing.




```
multicast-routing
 address-family ipv6
 interface all enable
```

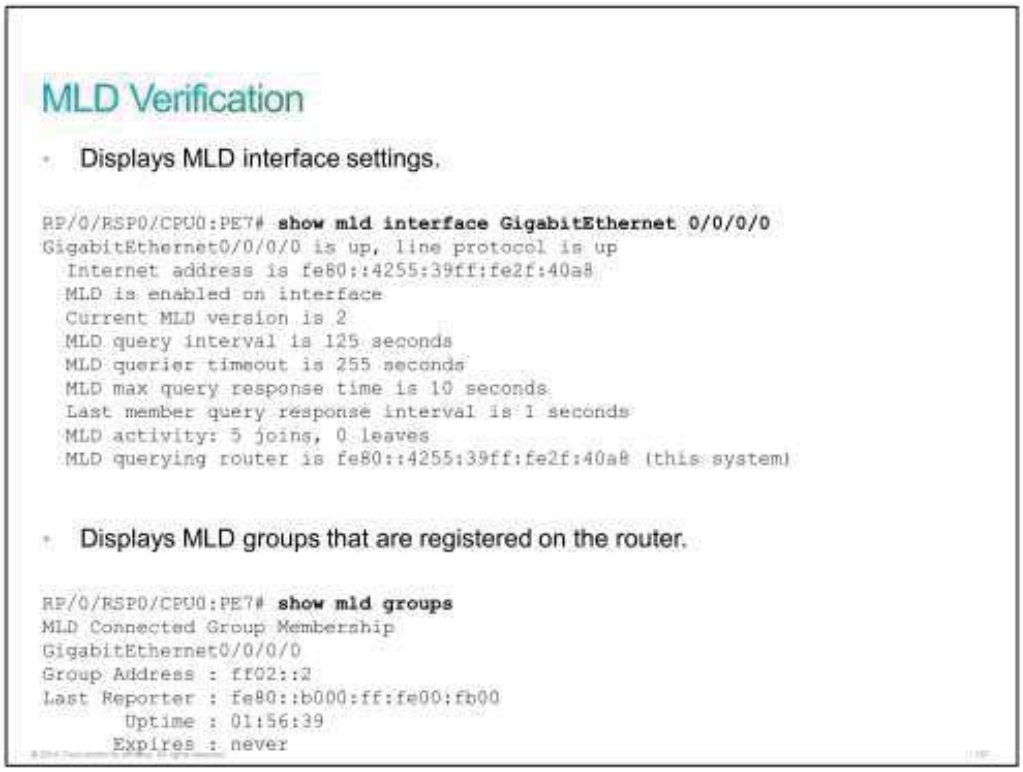
You can configure MLD in a similar way as IGMP. To configure the frequency at which Cisco IOS, IOS XE, IOS XR Software sends MLD host query messages, use the Cisco IOS and IOS XE Software **ipv6 mld query-interval** interface command or Cisco IOS XR Software **query-interval** router MLD command (or use the same command under a particular MLD interface). Similarly, you can configure the MLD version using the **version** command.



https://t.me/learningnets

# MLD Verification

To display MLD interface settings, use the Cisco IOS and IOS XE **show ipv6 mld interface** or Cisco IOS XR **show mld interface** command. To view the MLD groups that are registered on the router, use the Cisco IOS and IOS XE **show ipv6 mld groups** command or Cisco IOS XR **show mld groups** command.



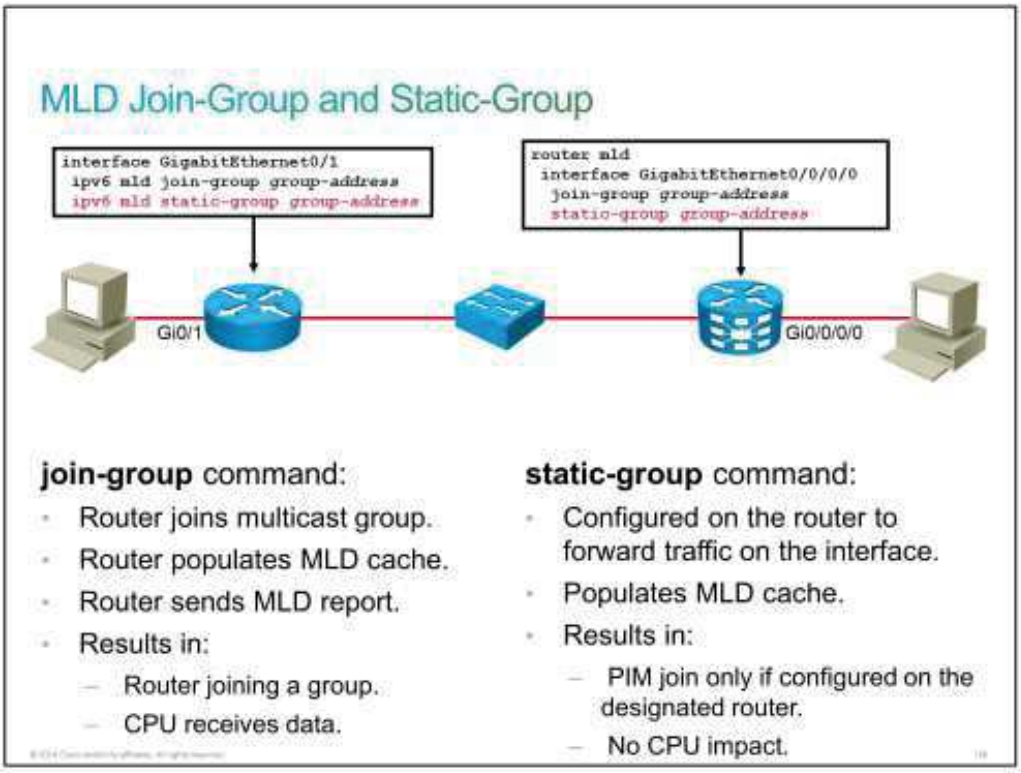

The figure shows two MLD verification outputs.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# MLD Join-Group and Static-Group

To have the router join a multicast group, use the Cisco IOS and IOS XE **ipv6 mld join-group** interface configuration command or Cisco IOS XR **join-group** command under the MLD configuration mode. This command causes the router to join a group and start listening to the multicast traffic for this group. This command makes the router CPU process the received multicast packets. This command causes the router to act like a multicast receiver for testing and troubleshooting purposes.



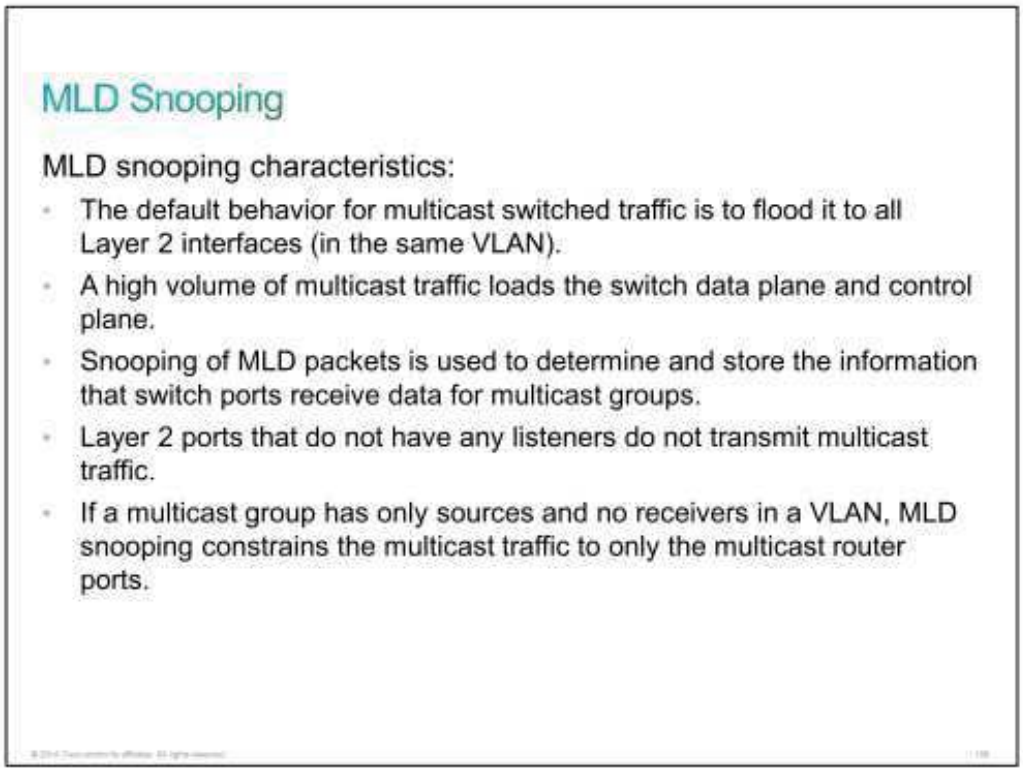

To configure the router to be a statically connected member of the specified group on the interface, use the Cisco IOS and IOS XE **ipv6 mld static-group** interface configuration command or Cisco IOS XR **static-group** command under the MLD configuration mode. If the router on which this command is configured is the designated router, it will generate a PIM join. This command does not have any effect on the router CPU because it does not process the multicast packets. This command causes the router to forward multicast traffic down an interface.



https://t.me/learningnets

# MLD Snooping

By default, Layer 2 switches flood all multicast traffic out of all interfaces in the same VLAN. Because a high volume of traffic can overload the switch data and control plane, mechanisms exist for how to limit traffic to only ports that have multicast listeners that are connected to them.




In IPv4, Layer 2 switches can use IGMP snooping to limit the flooding of multicast traffic by dynamically configuring Layer 2 interfaces so that multicast traffic is forwarded to only those interfaces that are associated with IP multicast devices. In IPv6, MLD snooping performs a similar function. With MLD snooping, IPv6 multicast data is selectively forwarded to a list of ports that want to receive the data, instead of being flooded to all ports in a VLAN. This list is constructed by snooping IPv6 multicast control packets.

MLD normal behavior:

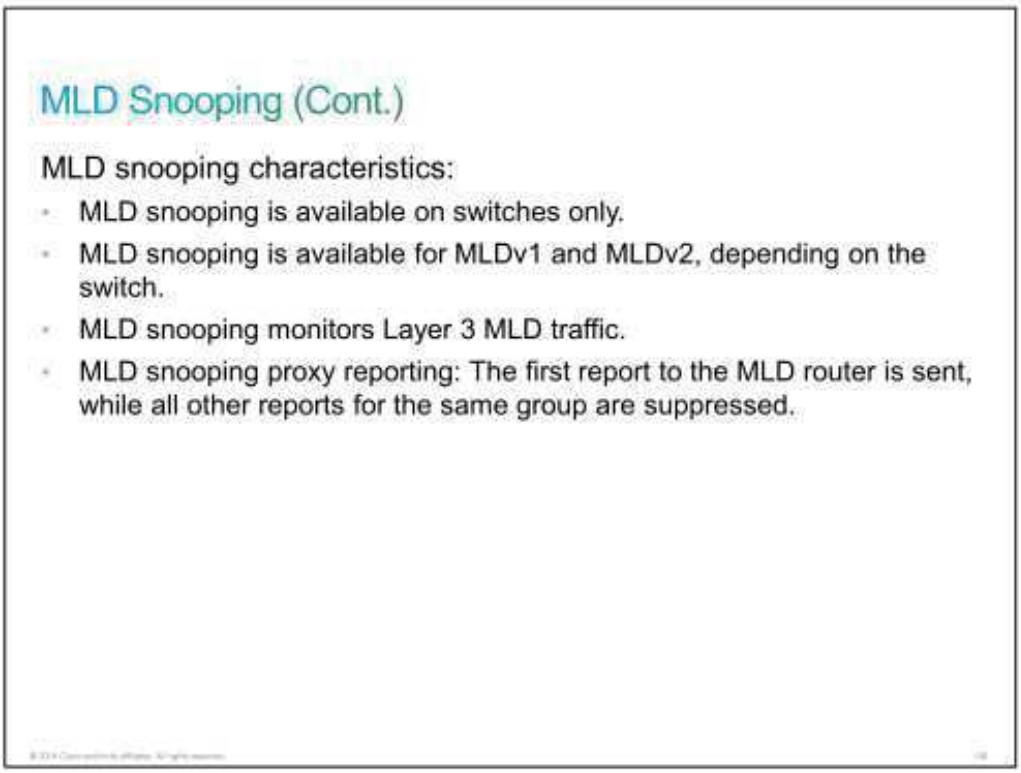
- MLD, which runs at Layer 3 on a multicast router, generates Layer 3 MLD queries in subnets where the multicast traffic needs to be routed.
- MLD (on a multicast router) sends out periodic general MLD queries that the switch forwards through all ports in the VLAN, and to which hosts respond. MLD snooping monitors the Layer 3 MLD traffic.

You can configure the switch to use MLD snooping in subnets that receive MLD queries from the MLD querier. MLD snooping constrains IPv6 multicast traffic at Layer 2 by configuring Layer 2 LAN ports dynamically to forward IPv6 multicast traffic only to those ports that want to receive it.

MLD is derived from IGMP; MLDv1 is equivalent to IGMPv2, and MLDv2 is equivalent to IGMPv3. MLD is a subprotocol of ICMPv6, and MLD messages are a subset of ICMPv6 messages that are identified in IPv6 packets by a preceding Next Header value of 58.



https://t.me/learningnets




A switch supports two versions of MLD snooping:

- MLDv1 snooping detects MLDv1 control packets and sets up traffic bridging that is based on IPv6 destination multicast addresses.
- MLDv2 basic snooping uses MLDv2 control packets to set up traffic forwarding that is based on IPv6 destination multicast addresses.

The switch can snoop on both MLDv1 and MLDv2 protocol packets and bridge IPv6 multicast data that is based on destination IPv6 multicast addresses.

Because MLD does not have report suppression, all of the hosts send their complete multicast group membership information to the multicast router in response to queries. The switch snoops these responses, updates the database, and forwards the reports to the multicast router. To prevent the multicast router from becoming overloaded with reports, MLD snooping does proxy reporting.

Proxy reporting forwards only the first report for a multicast group to the router and suppresses all other reports for the same multicast group.

Proxy reporting processes solicited and unsolicited reports. Proxy reporting is enabled by default and cannot be disabled.

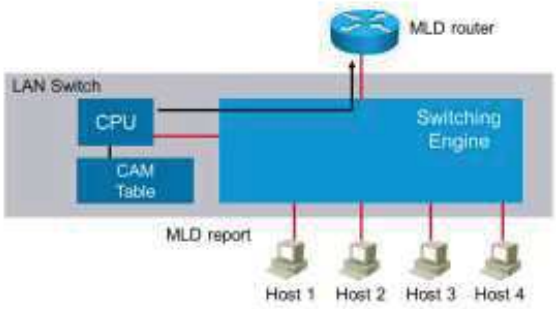


https://t.me/learningnets

## MLD Snooping (Cont.)

MLD snooping group join characteristics:

- Hosts join a multicast group.
- Switch creates a Layer 2 FIB entry for the multicast group in this VLAN.
- When additional hosts want to join the group, the switch adds its port to the Layer 2 FIB for the group.




Hosts join IPv6 multicast groups either by sending an unsolicited MLD report or by sending an MLD report in response to a general query from an IPv6 multicast router. (The switch forwards general queries from IPv6 multicast routers to all ports in a VLAN.) The switch snoops these reports.

In response to a snooped MLD report, the switch creates an entry in its Layer 2 forwarding table for the VLAN on which the report was received. When other hosts that are interested in this multicast traffic send MLD reports, the switch snoops their reports and adds them to the existing Layer 2 forwarding table entry. The switch creates only one entry per VLAN in the Layer 2 forwarding table for each multicast group for which it snoops an MLD report.

MLD snooping suppresses all but one of the host reports per multicast group and forwards this one report to the IPv6 multicast router. The multicast router begins sending the multicast group data to the switch.

The switch hardware can distinguish MLD information packets from other packets for the multicast group. Only MLD packets should be sent to the CPU, which prevents the switch from becoming overloaded with multicast frames. Frames that are addressed to the multicast MAC address (derived from the group address) that are not MLD packets should be sent to the multicast router and to the host that has joined the group.

If another host sends an unsolicited MLD report for the same group, the switch snoops this message and adds the port number of this host to the forwarding table.

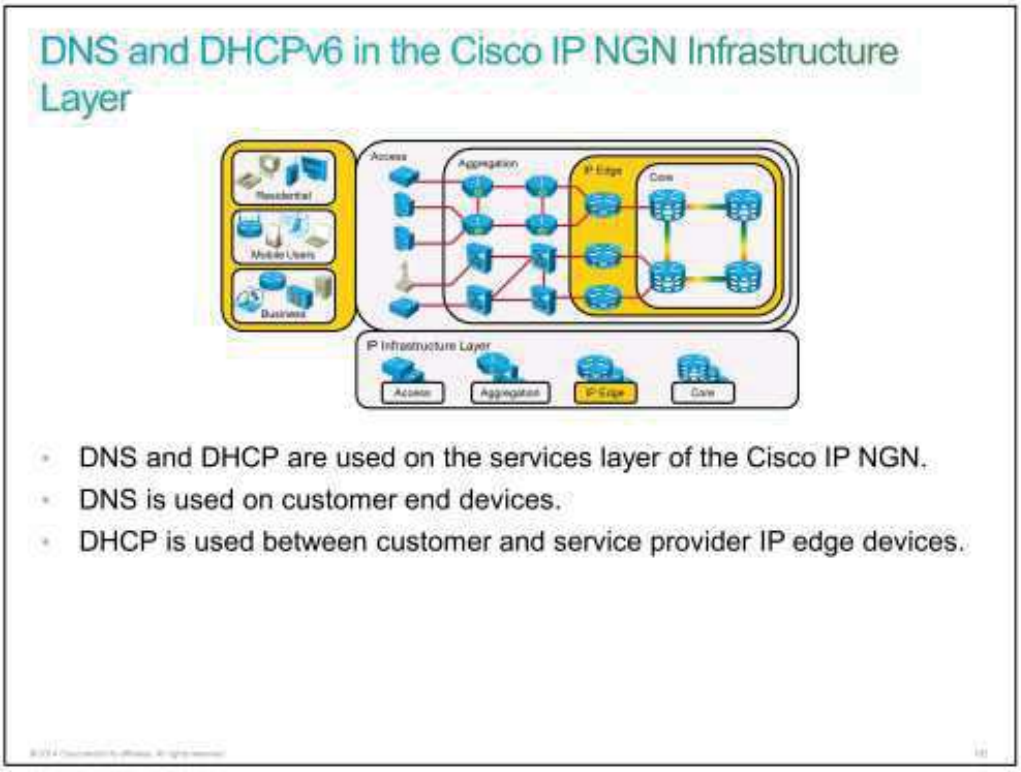
| | |
|---|---|
| **Note** | Because the forwarding table directs MLD messages only to the switch, the message is not flooded to other ports. Any known multicast traffic is forwarded to the group and not to the switch. |

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# DNS and DHCPv6 in the Cisco IP NGN Infrastructure Layer

DNS and DHCPv6 are used in the services layer of the Cisco IP NGN.




DNS is used by customer end devices. DHCP is used between CE and PE devices.



https://t.me/learningnets

# DNS IPv6 Support

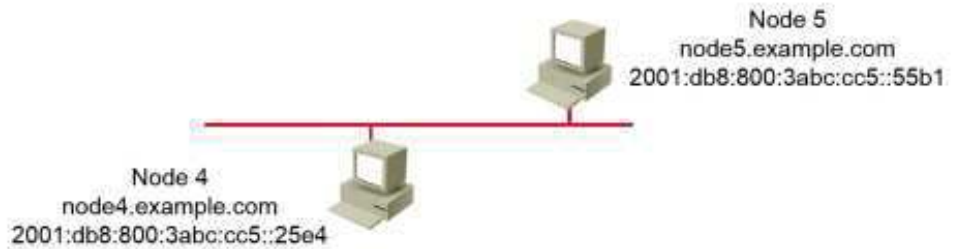
The DNS protocol had to be updated to support IPv6 in addition to IPv4. The two main tasks were to enable name lookup for IPv6 addresses and to enable the servers to communicate between themselves on IPv6 in addition to IPv4.

## DNS IPv6 Support

DNS Supported Objects:

- Two DNS issues exist for IPv6:
  - IPv6 record support.
  - IPv6 transport support.
- Several types of DNS objects exist:
  - AAAA, A, PTR, MX, etc.
- Forward lookups:
  - DNS uses AAAA records for forward IPv6 lookups.
  - PTR records are used for reverse lookups.

## DNS IPv6 Support (Cont.)




Examples of AAAA and A records:

```
node5.example.com.    IN  AAAA     2001:db8:800:3abc:cc5::55b1
node5.example.com.    IN    A        193.77.119.33
```

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

The DNS servers maintain a database for holding the relations between domain names (such as http://www.example.com) and IP addresses. This information is stored in DNS databases in the form of *records*. Depending on the record type (such as AAAA, A, PTR, MX, and so on), different information is stored. An MX record, for example, stores the IP address of the mail server for that domain (for example, http://mail.example.com).

Two types of lookups are used most in DNS:

- **Forward lookups**: provide resolution from a domain name to an IP or IPv6 address.
- **Reverse lookups**: provide resolution from an IP address to a domain name.

There are several types of objects in a DNS record about a domain. These include several types of records, as follows:

- **A records:** for IPv4 name-to-address lookups.
- **AAAA records:** for IPv6 name-to-address lookups.
- **MX records:** for the IP address of the mail server.

To support IPv6 in DNS, make these two updates to the DNS client and server systems:

- Update the DNS server and client to accept IPv6 record formats.
- Update the DNS server and client to run over IPv6 transport

These updates do not have to happen at the same time. Early DNS implementations often support the new records but run only over IPv4 transport. These early systems will work only for dual-stack clients and servers. An IPv6-only implementation would not work because DNS would not use IPv6 transport.

The three records or formats for IPv6 follow:

- Forward lookups
- Nibble format (reverse lookups)
- Bitwise format (reverse lookups)—deprecated

Bitwise format is no longer recommended and has been moved to experimental status, but some implementations still deploy it.

Forward lookups (name to IPv6 address) are completed via the AAAA record, which is the address record for IPv6 DNS. This record links a hostname to a 128-bit address, which is the forward lookup record.

Here is an example of a AAAA record:

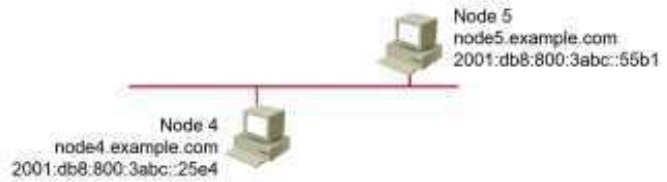
```
$ORIGIN example.com.
node4 3600 IN AAAA 2001:db8:800:3abc:cc5::25e4
node5 3600 IN AAAA 2001:db8:800:3abc:cc5::55b1
```

| **Note** | There were A6 records to resolve an IPv6 address from a name; however, they are deprecated. AAAA records are used instead. |
|---|---|



https://t.me/learningnets

## DNS IPv6 Support (Cont.)




Reverse lookups:

- IPv6 uses PTRs for reverse lookups, similar to IPv4, but with the new nibble format.
- Examples of Nibble-Formatted Records:

```
$ORIGIN c.b.a.3.0.0.8.0.8.b.d.0.1.0.0.2.ip6.arpa.
4.e.5.2.0.0.0.0.0.0.0.0.0.0.0.0 14400 IN PTR node4.example.com.
1.b.5.5.0.0.0.0.0.0.0.0.0.0.0.0 14400 IN PTR node5.example.com.
```

Reverse lookups (address to name) are still accomplished using the PTR. There are two formats for address representation: one recommended and one deprecated.

The nibble format is preferred. It uses the top-level domain "ip6.arpa." (Initially, the top-level domain was called "ip6.int," but this convention was deprecated in RFC 4159 and does not need to be maintained any longer.) Notice in the following example that address representation is backward, with each 4-bit position (one hexadecimal character) separated by a "." (dot). There is no compressed format for the address, so you cannot eliminate leading zeros.

```
$ORIGIN c.b.a.3.0.0.8.0.8.b.d.0.1.0.0.2.ip6.arpa.
4.e.5.2.0.0.8.0.8.b.d.0.1.0.0.2 14400 IN PTR node4.example.com.
1.b.5.5.0.0.8.0.8.b.d.0.1.0.0.2 14400 IN PTR node5.example.com.
```

 © 2014 Cisco Systems, Inc.

https://t.me/learningnets

## DNS IPv6 Support (Cont.)

DNS tree structure characteristics:

- IPv6 needs an updated version of a DNS server and client resolver.
- DNS tree structure is identical to IPv4:
  - Root DNS server.
  - Top-level domain DNS server.
  - Authoritative DNS server for each particular domain.
- From the operational perspective, there are:
  - Authoritative DNS servers.
  - Caching DNS servers.
- The majority of DNS root servers are accessible using IPv6, many since 2008:
  - Enabled, end-to-end IPv6 communication without using IPv4 for communication with the root DNS server.
  - Removed the need for dual stack (from DNS perspective).

The hierarchy of DNS servers is best described with a tree. On the top of the hierarchy, there are root DNS servers (and only 13 clusters of these servers in the world).

Below the root server are TLD DNS servers that resolve IP addresses for TLDs, such as .com, .net, .org, .us, .uk, and so on.

Beneath TLD servers, there are authoritative DNS servers for each domain. These resolve IP addresses from their domains only (such as for http://example.com).

The IPv6 DNS tree structure is identical to the deployed structure for IPv4. Clients query local caching servers, which locate the DNS server with the authoritative records for a given zone through message exchange with a root DNS server. These caching servers then return records to the client (and cache the information locally for near-term future use). Typically, a protocol-independent application will query first for an AAAA record, and then, if the DNS returns "no such records," will query for an A record. Multiple records may be returned.

These are the major components of the DNS tree structure:

- Root DNS
- Authoritative DNS
- Caching DNS (typically also deployed in sets, not a single machine)
- Client-based DNS resolver library

For redundancy and operational efficiency reasons, there are primary and secondary DNS servers for every hierarchy level, and there are cache DNS servers that cache results of DNS queries within enterprise networks.

Root DNS servers contain records that link domain names to their authoritative DNS servers.



https://t.me/learningnets

The records that are maintained by the root DNS server for a given domain name should include at least one IPv4 address for the authoritative server (for a given zone). The records may contain more than one IPv4 address and also may contain multiple IPv6 addresses. This inclusion prevents situations in which an IPv4-only caching server is referred to as an IPv6-only authoritative server, to which it could clearly not connect.

An example record (in generalized format) would be the following:

```
"example.com - Authoritative Primary DNS at 2001:db8:400::200c, Secondary at
2001:db8:100::4e20"
"example.com - Authoritative Primary DNS at 192.0.2.10, Secondary at 192.0.2.20"
"2001:db8:800:3abc:cc5::25e4" - Authoritative Primary (2001:db8:800/48) at
2001.db8:700:abcd::1000,
Secondary at 2001:db8:600:ef12::2000
"2001:db8:800:3abc:cc5::25e4" - Authoritative Primary (2001:db8:800/48) at
192.0.2.130,
Secondary at 192.0.2.140
```

The fact that the root DNS servers only advertise their IPv4 addresses (even though some do respond on IPv6) means that it was not possible to deploy an IPv6-only enterprise. Within the enterprise, all DNS servers can be implemented using IPv6 transport: "Resolver client to local caching server."

The problem is that the caching server had to talk to the root DNS server over IPv4 transport, so it had to be a dual-stack node, not an IPv6-only node.

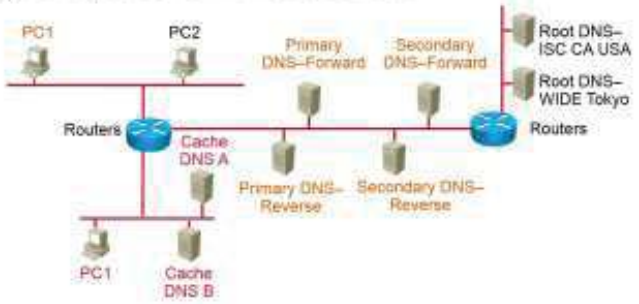
Furthermore, many of the ccTLD DNS servers (that is, servers providing a domain name for a specific country code, such as ".us") were also IPv4-only.

| **Note** | IANA added AAAA records for its root DNS servers in 2008. Since then, approximately half of the servers are reachable using IPv6, making IPv6-only networks possible. |
|---|---|

## DNS IPv6 Support (Cont.)

DNS tree structure components characteristics:

- Authoritative primary and secondary DNS servers support both IPv6 and IPv4 records:
  - Forward and reverse zones are not often on the same system.
  - Reverse zones are often maintained by an ISP.
- Caching DNS is typically provided by ISPs (for home or small business) or by large enterprises for in-house clients.




Examples of records that are maintained on these DNS servers are as follows:



https://t.me/learningnets

These are DNS tree structure components:

- **Root DNS:** Root DNS servers contain records that link domain names to their authoritative DNS servers. There are currently 13 root DNS IP addresses. (There are more than 13 servers. Many addresses are IPv4-anycast addresses and "front" a number of servers.) The root DNS servers are not uniformly addressable on their IPv6 addresses; some are reachable over IPv6 transport, but several still are not reachable.
- **Top-Level DNS:** Top-Level DNS servers resolve IP addresses for TLDs, such as .com, .net, .org, .info, .biz, and for ccTLDs, such as .us, .uk, .de, .hk, .au, and so on.
- **Authoritative Primary DNS:** For a given domain, authoritative primary DNS servers contain the official records for hosts within a given domain name. For reverse lookups, authoritative primary DNS servers contain the official reverse-lookup records for the given IP address. Typically, the forward authoritative DNS server is not the same host as the secondary authoritative DNS server.

```
"node4.example.com" - 2001:db8:800:3abc:cc5::25e4
"2001:db8:800:3abc:cc5::25e4" - node4.example.com
```

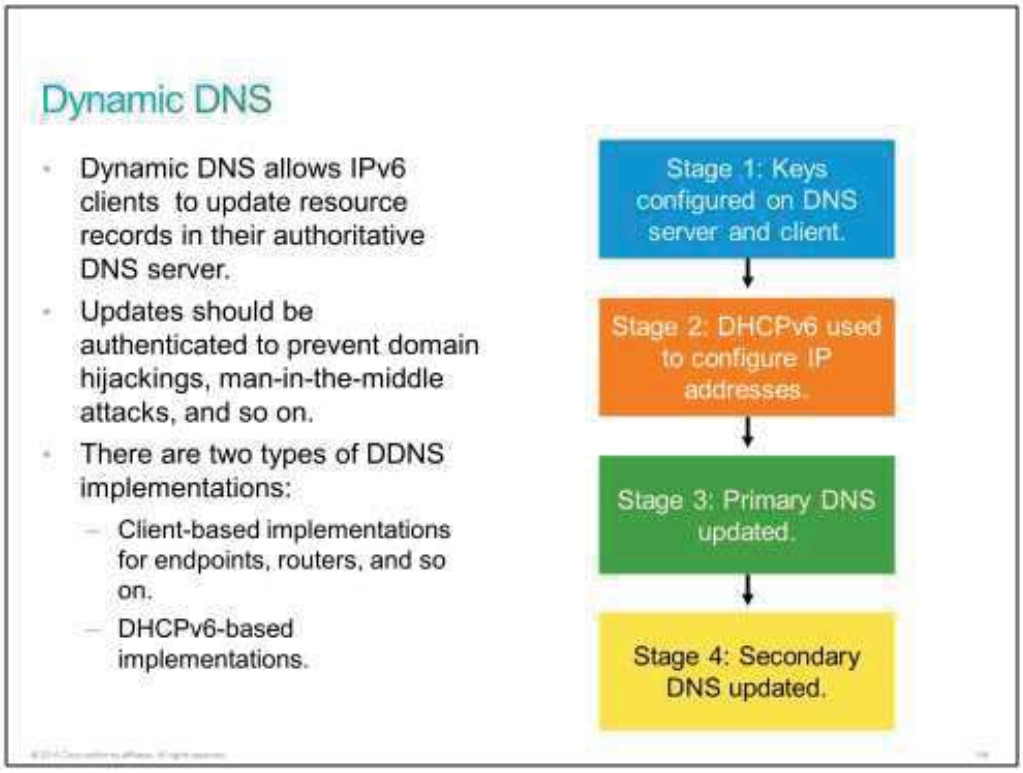
- **Secondary DNS:** For a given domain, secondary DNS servers provide a backup if the primary DNS server fails. Secondary DNS servers periodically transfer records from the primary DNS server.
- **Caching DNS:** Caching DNS servers answer queries from client devices and help reduce the load on the primary, secondary, and root DNS servers. No records are permanently maintained on caching DNS servers. When a caching DNS server helps resolve a record on behalf of a client, it stores the record locally in a cache for a time—to use when answering other clients that are asking for the same record—before discarding it.
- **Client Devices:** Client devices are IP nodes that use a DNS resolver to translate names to addresses and addresses to names. Client devices are configured to point to multiple caching servers.



https://t.me/learningnets

# Dynamic DNS

Historically, IP-based servers had their addresses manually configured into the primary DNS server. These addresses were also statically assigned on the node; therefore, the addresses and name-to-address translation was long-lived. For example, hostname "media.example.com" would be at IP address 192.168.0.200, and this entry would be in the primary (and secondary) authoritative DNS server. Client machines usually did not have an entry in DNS because the ability to be reliably contacted by a peer node was not desired or practical.




When most devices are both clients and servers (in other words, peers—an important driver for IPv6 adoption), and those devices configure their addresses via, for example, DHCPv6 or autoconfiguration, those nodes need to dynamically create or update their DNS records on their authoritative primary DNS server. This process gives the nodes stable host-to-address and address-to-host mappings even when their dynamically assigned addresses change. Using DDNS, DHCPv6 clients can dynamically update their records in DNS. DDNS is still under active discussion in the IETF working groups. Many published RFCs and drafts that are related to DDNS are in circulation.

The DDNS process goes through these stages:

- **Stage 1:** Keys are configured on the DNS server and client. DDNS exchanges must be secured; otherwise, man-in-the-middle attacks, in which a malicious party captures traffic that is intended for another node, are possible.
- **Stage 2:** The IPv6 node uses DHCPv6 to configure an IP address or other information. The address also can be configured via stateless autoconfiguration, and the DDNS update can be performed in the same manner.
- **Stage 3:** The primary DNS is updated. The DHCPv6 client on the node updates the primary DNS server for both forward and reverse records.



https://t.me/learningnets

- **Stage 4:** The secondary DNS is updated. The primary DNS server updates the secondary DNS server via zone transfer.

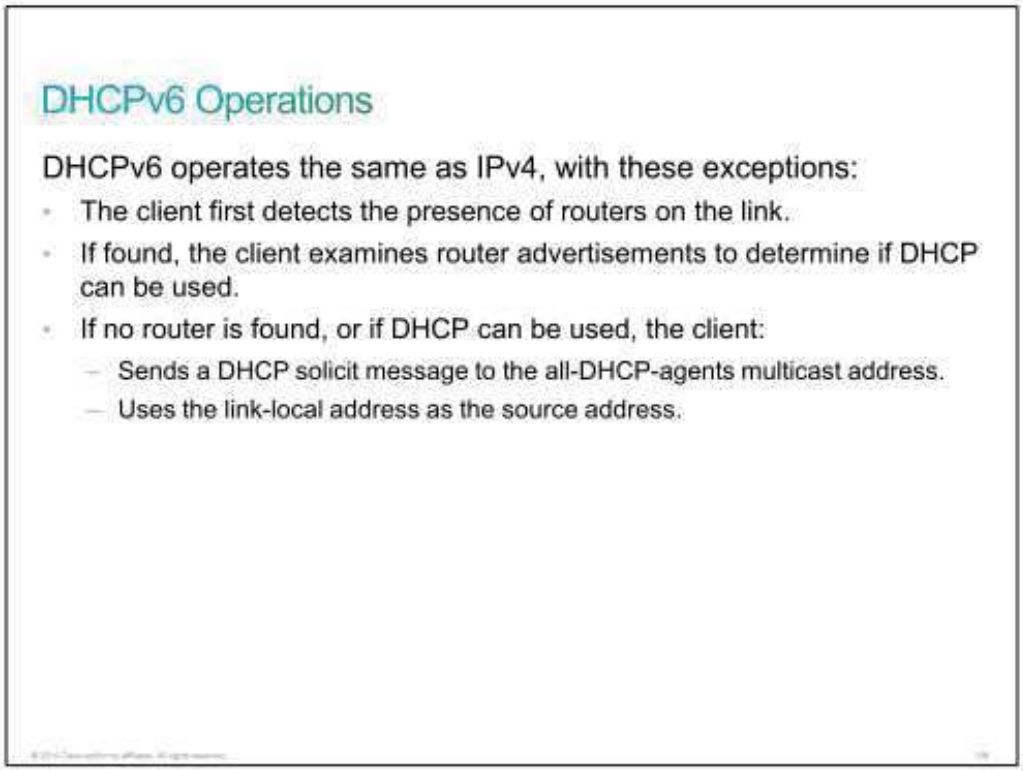
| | |
|---|---|
| **Note** | DDNS is an important building block for IPv6, even though it is not strictly related to IPv6. In IPv4, most network applications are client-server based, such as web servers (servers) and browsers (clients). In IPv4 application architecture, clients are normally anonymous—they have no entry in the global DNS system—and servers are only reachable at well-known DNS names. |
| **Note** | One compelling IPv6 feature is the ability to support peer computing, where the terms "client" and "server" are no longer meaningful. All nodes are complete peers on the network and reachable via their well-known DNS name. This scenario implies that all nodes have current entries in the DNS. From a scalability perspective, and considering that many nodes will be mobile and will use autoconfiguration or DHCPv6 for address assignment, DDNS is the only reasonable solution to ensure that these nodes always have a current DNS mapping. |



https://t.me/learningnets

# DHCPv6 Operations

Acquiring configuration data for a client in DHCPv6 is like the process in IPv4 but with a few exceptions.




The client can sometimes detect the presence of routers on the link using neighbor discovery messages. If at least one router is found, the client examines the router advertisements to determine if DHCP should be used. If the router advertisements allow use of DHCP on this link, or if no router is found, the client starts a DHCP solicit phase to find a DHCP server.

DHCPv6 uses multicast for many messages. When the client sends a solicit message, it sends the message to the all-DHCP-agents multicast address with link-local scope. Agents include both servers and relays.

When a DHCP relay forwards a message, the relay can forward the message to the all-DHCP-servers multicast address with site-local scope. This means that a relay does not need to be configured with all of the static addresses of the DHCP servers, as in IPv4. If needed by policy, a relay can contain a static list of DHCP servers.

Some servers can be configured to give global addresses using policies, for example, "do not give to printer." Other servers (or the same servers within a different context) can be configured to give site-local addresses using a different policy, for example, "give to anyone."

| | |
|---|---|
| **Note** | DHCPv6 solicit messages are sent from the link-local address of the requesting node, which is the address that the node constructs for itself at initialization. The request is sent to a reserved DHCPv6-specific multicast address. This process differs markedly from the IPv4 practice, where the message is sent from the unspecified address to the broadcast address. This scenario is an excellent example of IPv6 using more elegant mechanisms than IPv4 to improve network scalability. |



https://t.me/learningnets

## DHCPv6 Operations (Cont.)

DHCPv6 operates using the following multicast adresses:

| IPv6 Multicast Address | Description |
|---|---|
| FF02::1:2 | All-DHCP-agents (servers or relays), link-local scope |
| FF05::1:3 | All-DHCP-servers, site-local scope |
| FF05::1:4 | All-DHCP-relays, site-local scope |

DHCPv6 uses these multicast addresses:

- FF02::1:2 is the all-DHCP-agents (servers or relays) address that clients use to communicate with unknown agents on their local link. This address is of link-local scope.
- FF05::1:3 is the all-DHCP-servers address that is used by clients to communicate with unknown site-wide servers. This address is of site-local scope. A node that sends a message to this destination address should use a source address that will be reachable by the server for the answer. For example, a link-local address cannot be used as a source address for this kind of message.
- FF05::1:4 is the all-DHCP-relays address. Because it is site-local address, it can be used to reach all DHCP relays within one site.

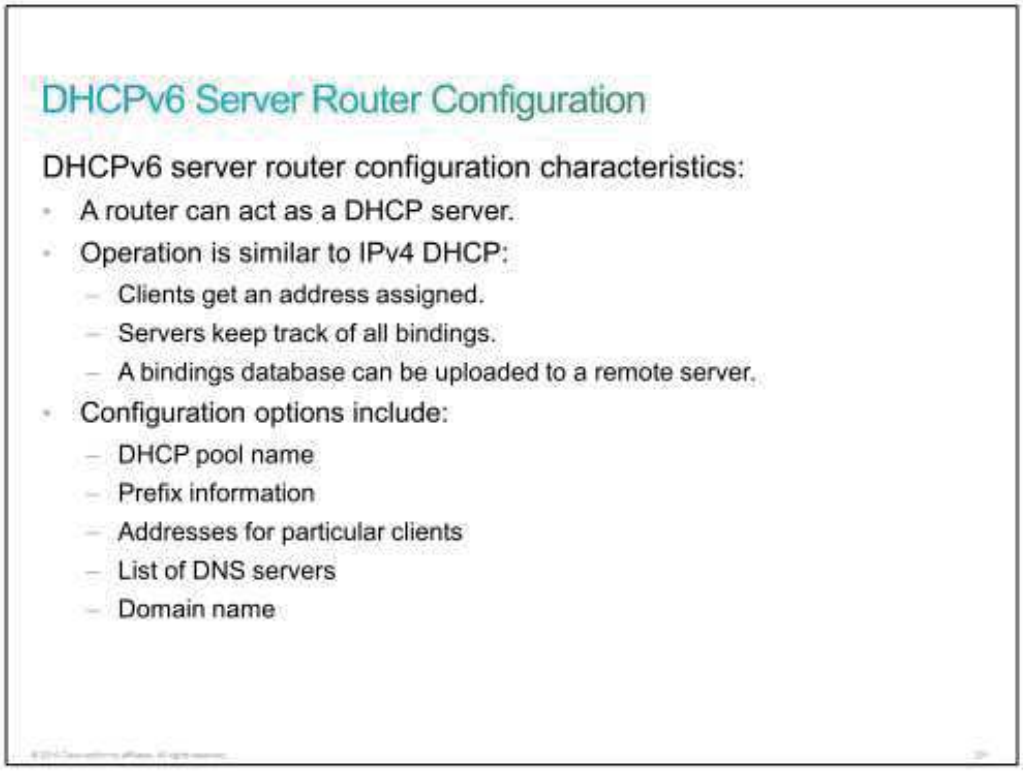
Note that DHCPv6 servers and relay agents listen on UDP port 547, while DHCPv6 clients listen on UDP port 546.



https://t.me/learningnets

# DHCPv6 Server Router Configuration

The DHCPv6 server router configuration is described below:




The DHCPv6 server function can be enabled on individual IPv6-enabled interfaces.

The DHCPv6 server can provide those configuration parameters that do not require the server to maintain any dynamic state for individual clients, such as DNS server addresses and domain search list options. The DHCPv6 server may be configured to perform prefix delegation.

All of the configuration parameters for clients are independently configured into DHCPv6 configuration pools, which are stored in NVRAM. A configuration pool can be associated with a particular DHCPv6 server on an interface when it is started. Prefixes to be delegated to clients may be specified either as a list of preassigned prefixes for a particular client or as IPv6 local prefix pools, which are also stored in NVRAM. DHCPv6 configuration pools can reference and use the list of manually configured prefixes or IPv6 local prefix pools.

The DHCPv6 server maintains an automatic binding table in memory to track the assignment of some configuration parameters, such as prefixes between the server and its clients. The automatic bindings can be stored permanently in the database agent, which can be, for example, a remote TFTP server or local NVRAM file system.

A DHCPv6 configuration information pool is a named entity that includes information about available configuration parameters and policies that control assignment of the parameters to clients from the pool. A pool is configured independently of the DHCPv6 service and is associated with the DHCPv6 service through the CLI.

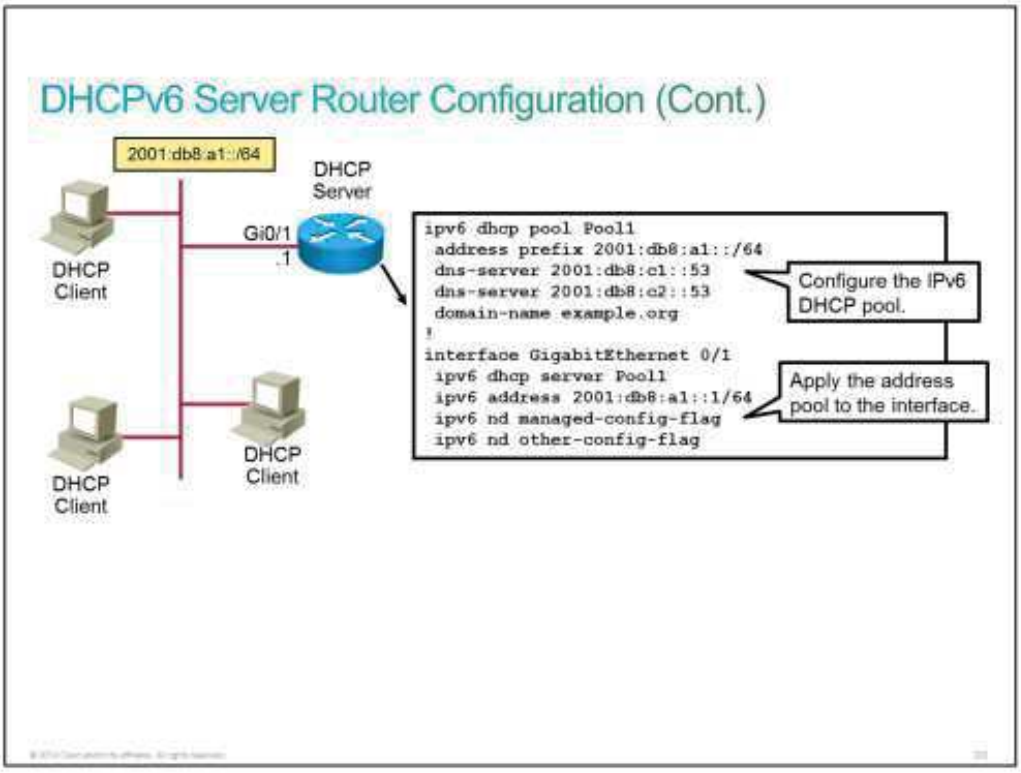
Each configuration pool can contain the following configuration parameters and operational information:

- Prefix delegation information that can include the following:
  - A prefix pool name and associated preferred and valid lifetimes.



https://t.me/learningnets

  - A list of available prefixes for a particular client and associated preferred and valid lifetimes.
- A list of IPv6 addresses of DNS servers.
- A domain name.




The figure provides an example of the DHCPv6 server configuration on Cisco IOS Software. First, configure the DHCP pool using the **ipv6 dhcp pool** command. Then specify the pool of addresses using the **address prefix** command and then specify the DNS server(s) using the **dns-server** command. Optionally, specify the domain name using the **domain-name** command. Alternatively, if you receive a domain name and DNS information from the upstream DHCPv6 server, you can import this information into your DHCPv6 server using the **import dns-server** and **import domain-name** commands.

Then enter the interface configuration mode for an interface that you would like to enable the DHCP server on. Then enable the DHCP server on the interface using the **ipv6 dhcp server** command, followed by a name of the configured DHCP pool.

It is also necessary to instruct a client to start a DHCP solicit phase to find a DHCP server that is using the **ipv6 nd managed-config-flag** command. This command sets the "managed address configuration highlight" in IPv6 router advertisements, which indicates to attached hosts that they should use DHCPv6 to obtain addresses. You can also configure the **ipv6 nd other-config-flag** command, which instructs the client to also obtain other parameters, such as DNS servers or a domain name, via DHCPv6.

The figure shows the configuration for the Cisco IOS and IOS XE Software platforms because the DHCPv6 server configuration for end clients is not supported on Cisco IOS XR Software platforms.

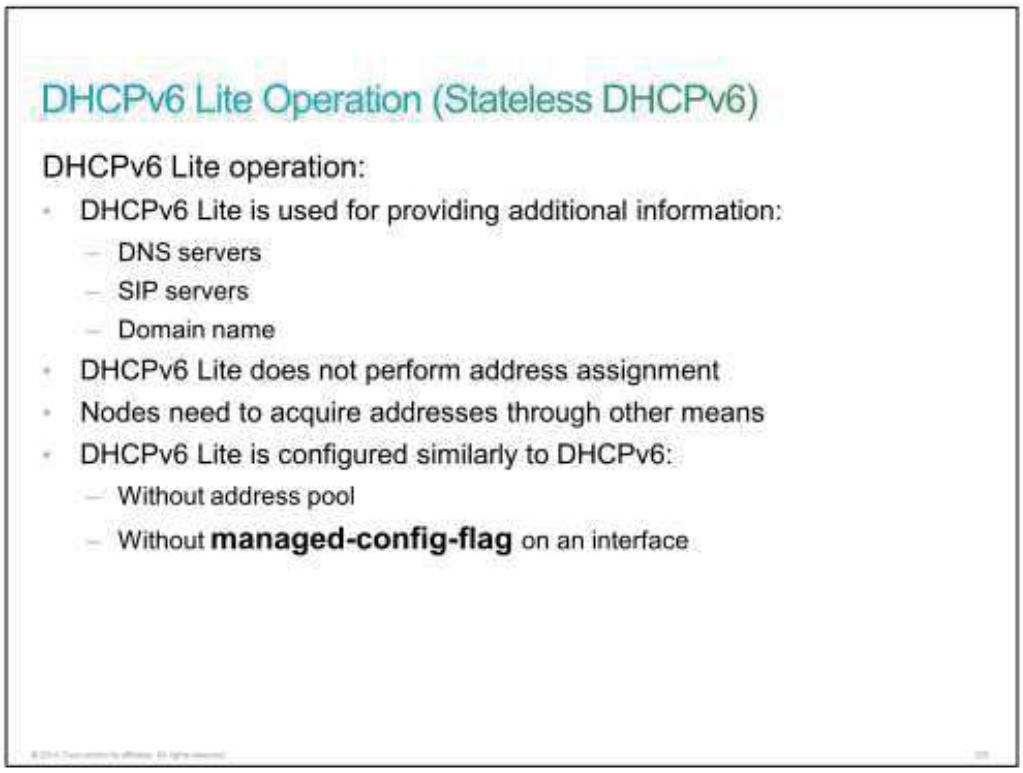
| **Note** | For a complete command reference, refer to the Cisco IOS and IOS XE Software Command Reference guides on www.cisco.com. |
|---|---|



https://t.me/learningnets

# DHCPv6 Lite Operation (Stateless DHCPv6)

DHCPv6 Lite (or stateless DHCPv6) is used in an environment where end nodes acquire IPv6 addresses through different means (most often using stateless autoconfiguration). However, these end nodes also need to obtain additional information (usually a list of DNS servers).



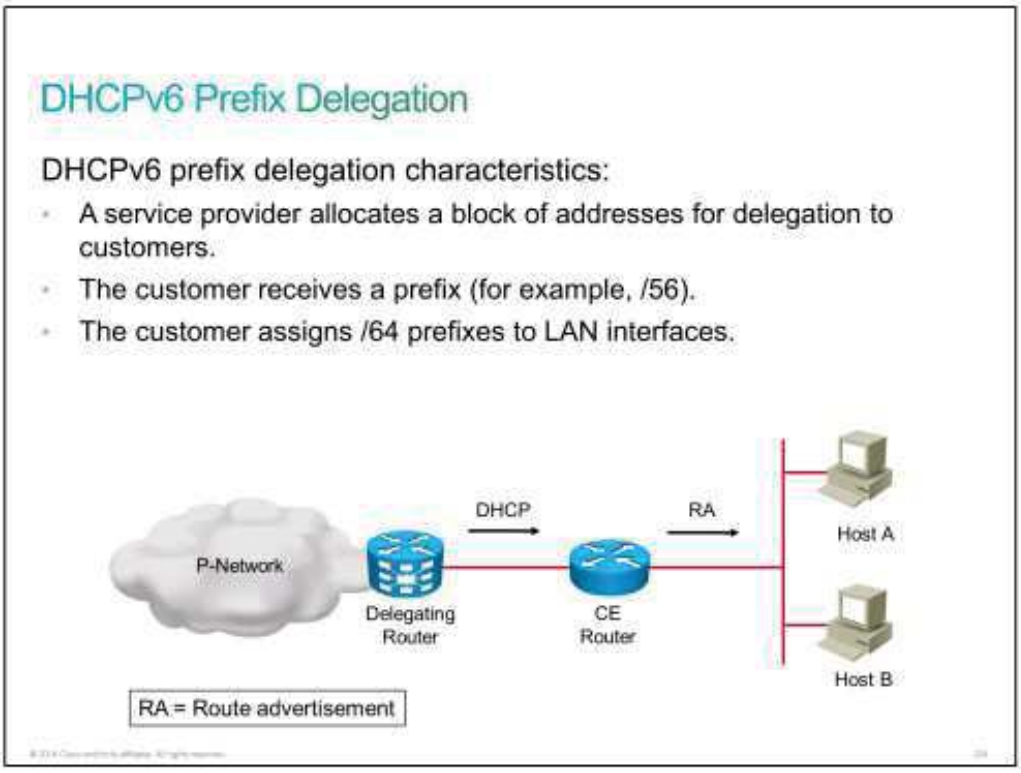

A stateless DHCPv6 server will send additional information if it is contacted by a stateless client. The client will be updated with additional information through DHCP Lite by using a route advertisement message. The configuration is similar to the DHCPv6 configuration that was described earlier. The configuration should be without an address pool and without the **ipv6 nd managed-config-flag** command on an interface.



https://t.me/learningnets

# DHCPv6 Prefix Delegation

Extensions to DHCPv6 enable prefix delegation, through which a service provider can automate the process of assigning prefixes to a customer for use within the customer network.




Prefix delegation occurs between a PE device and the CE, using the DHCPv6 prefix delegation option. After the service provider has delegated prefixes to a customer, the customer may further subnet and assign prefixes to the links in the customer network.

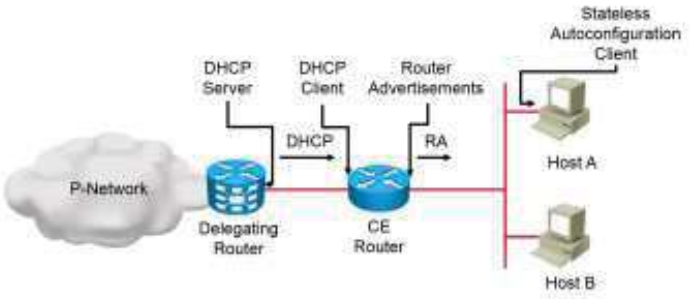


https://t.me/learningnets

## DHCPv6 Prefix Delegation (Cont.)

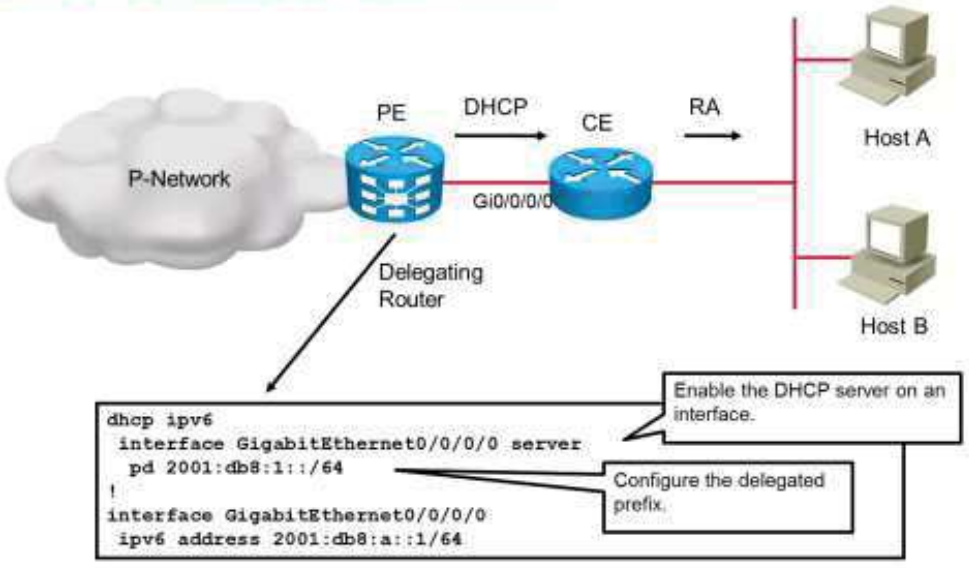
Interface configuration:

- PE as the delegating DHCP server.
- CE as the DHCP client and IPv6 router.




In the figure, the service provider-delegating router will act as a DHCP server and will allocate a prefix to the CE router. The CE will, on one side facing the delegating router, act as a DHCP client, acquire the prefix, and then assign smaller prefixes to its own local interfaces. On the CE interfaces that are facing the customer internal networks, the CE will act as an IPv6 router, sending out router advertisements to inform the local clients of prefix availability. In this configuration, the service provider indirectly assigns addresses to end nodes.

## DHCPv6 Prefix Delegation (Cont.)




 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

The figure shows the DHCPv6 prefix delegation configuration on the PE router. The router is running Cisco IOS XR Software.

| | |
|---|---|
| **Note** | As mentioned before, the Cisco IOS XR Software router cannot be configured as a DHCPv6 server for end clients. However, it can be configured as a DHCPv6 prefix delegation server. |

To enable DHCPv6 prefix delegation, first enter the DHCP configuration mode using the **dhcp ipv6** command. Then enable the DHCPv6 server on an interface that is facing the CE router using the **interface server** command. Then use the **pd** command, followed by a prefix that will be delegated to a client on the interface.

The configuration of the DHCPv6 prefix delegation prefix on Cisco IOS and IOS XE Software platforms is as follows:

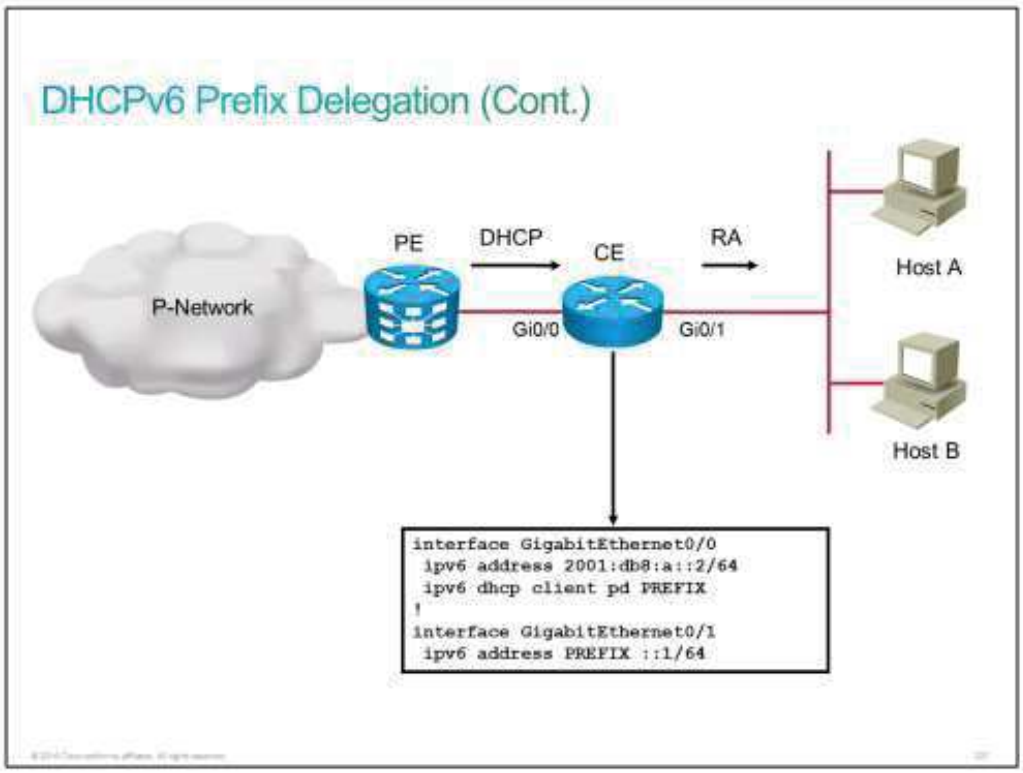
```
ipv6 dhcp pool Customers
prefix-delegation pool C_PREFIX
!
interface FastEthernet0/0
ipv6 address 2001:db8:a::1/64
ipv6 dhcp server Customers
!
ipv6 local pool C_PREFIX 2001:db8:c::/40 48
```

A DHCP pool that is named "Customers" has a **prefix-delegation pool** command with a reference to a local pool that is named "C_PREFIX." The local pool C_PREFIX contains specifications mandating that addresses will be allocated to clients along with prefix length. Addresses that are specified with 2001:db8:c::/40 are reserved for further allocation in blocks of /48. Therefore, each client will receive one /48 from the /40. The DHCPv6 server is enabled on the interface using the **ipv6 dhcp server** command, followed by a name of the configured DHCP pool.

| | |
|---|---|
| **Note** | The option to specify the prefix length for further allocation is available on Cisco IOS and IOS XE Software only. |
| **Note** | For a complete command reference, refer to the Cisco IOS, IOS XE, and IOS XR Software Command Reference guides on www.cisco.com. |



https://t.me/learningnets




The figure shows a sample configuration of a CE router. The interface that is facing the service provider acts as a client with the **ipv6 dhcp client pd** command, and the interface has a prefix delegation reference called "PREFIX." This prefix delegation will make it possible to refer to the allocated prefix with the variable PREFIX later on.

The interface that is facing the LAN devices is configured with an IPv6 address with a reference to the prefix name PREFIX. The service provider defined the first 64 bits (that is, the configuration on the delegating router in the previous slide), so only the last 64 bits must be defined. For stateless autoconfiguration to work, the network mask is set to /64.

| **Note** | For a complete command reference, refer to the Cisco IOS and IOS XE Software Command Reference guides on www.cisco.com. |
|---|---|

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# DHCPv6 Verification

The DHCPv6 verification commands on the router are described below:

## DHCPv6 Verification

DHCPv6 verification commands:

- Shows all DHCPv6 pools on a router.

```
Router# show dhcp ipv6 pool
```

- Shows the state of all current clients of the DHCP server.

```
Router# show dhcp ipv6 binding
```

## DHCPv6 Verification (Cont.)

- Displays whether an interface is in client mode or in server mode.

```
Router# show dhcp ipv6 interface
```

- Used when debugging either DHCP server or DHCP client functionality on a router.

```
Router# debug dhcp ipv6 detail
```



https://t.me/learningnets

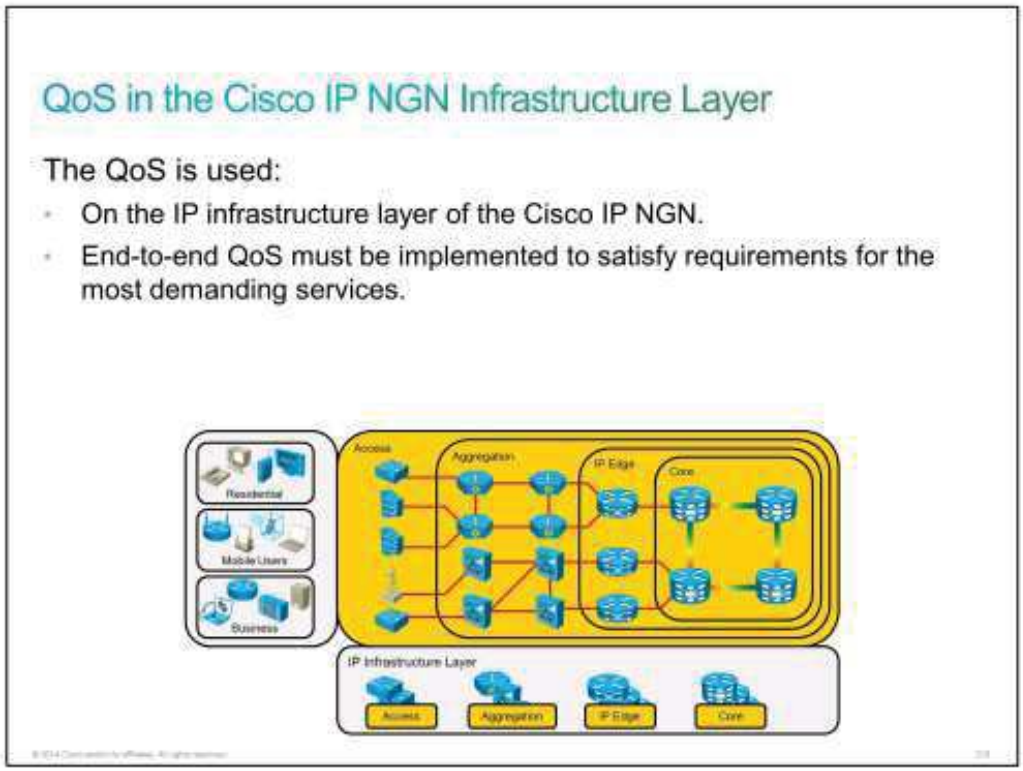
The figure presents various commands that are available to verify and troubleshoot DHCPv6 configuration and operations.

 © 2014 Cisco Systems, Inc.

https://t.me/learningnets

# QoS in the Cisco IP NGN Infrastructure Layer

QoS is used on the IP infrastructure layer of the Cisco IP NGN. In order to provide the highest level of service quality, QoS must be implemented across all areas of the network. Because of this, different QoS tools must be implemented in all parts of IP infrastructure layer.



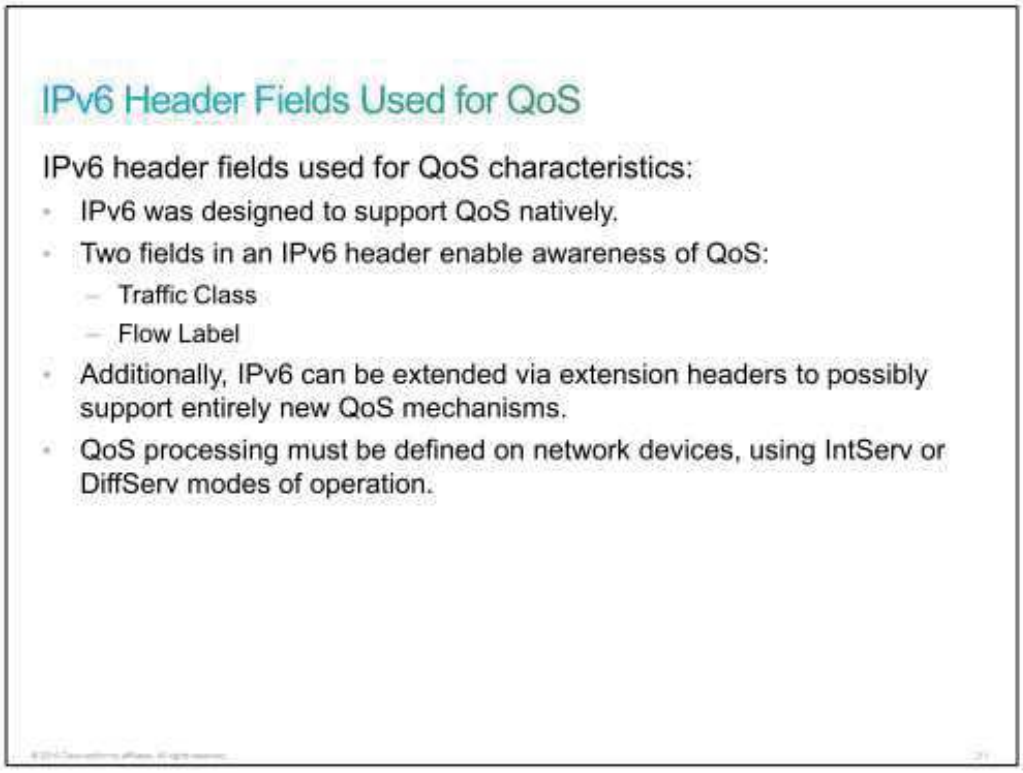

| Note | QoS is covered in detail in the *Implementing Cisco Service Provider Next-Generation Core Network Services* (SPCORE) v1.2 course. |
|---|---|



https://t.me/learningnets

# IPv6 Header Fields Used for QoS

The fields that are used in the IPv6 header to support QoS functions are described below:




IPv6 was designed to natively support QoS from the beginning. The IPv6 header contains two different fields that are designed to support QoS:

- Traffic Class (8 bits)
- Flow Label (20 bits)

In addition, because of the expanded reach of IPv6 via extension headers, you can add new features to IPv6 by defining new options to put into either the hop-by-hop options header or the destination options header. You can also create entirely new extension headers.

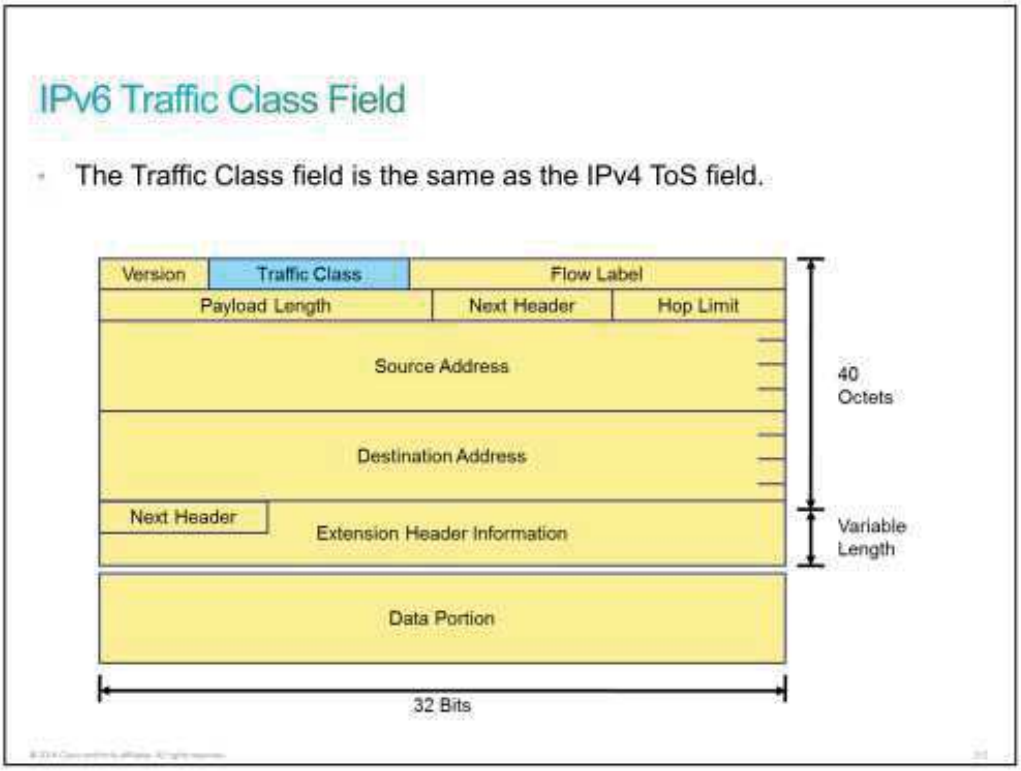
However, entirely new QoS paradigms and mechanisms are not really needed, so current QoS mechanisms are used. Current QoS implementations are based either on the IntServ QoS model or on the DiffServ QoS model. IntServ is used when absolute QoS guarantees are needed. DiffServ defines "soft" QoS guarantees by just prescribing the behavior of a device that is based on the priority of the packet (or the PHB). For example, there is absolute priority forwarding for important packets and normal forwarding operation for all other packets.



https://t.me/learningnets

# IPv6 Traffic Class Field

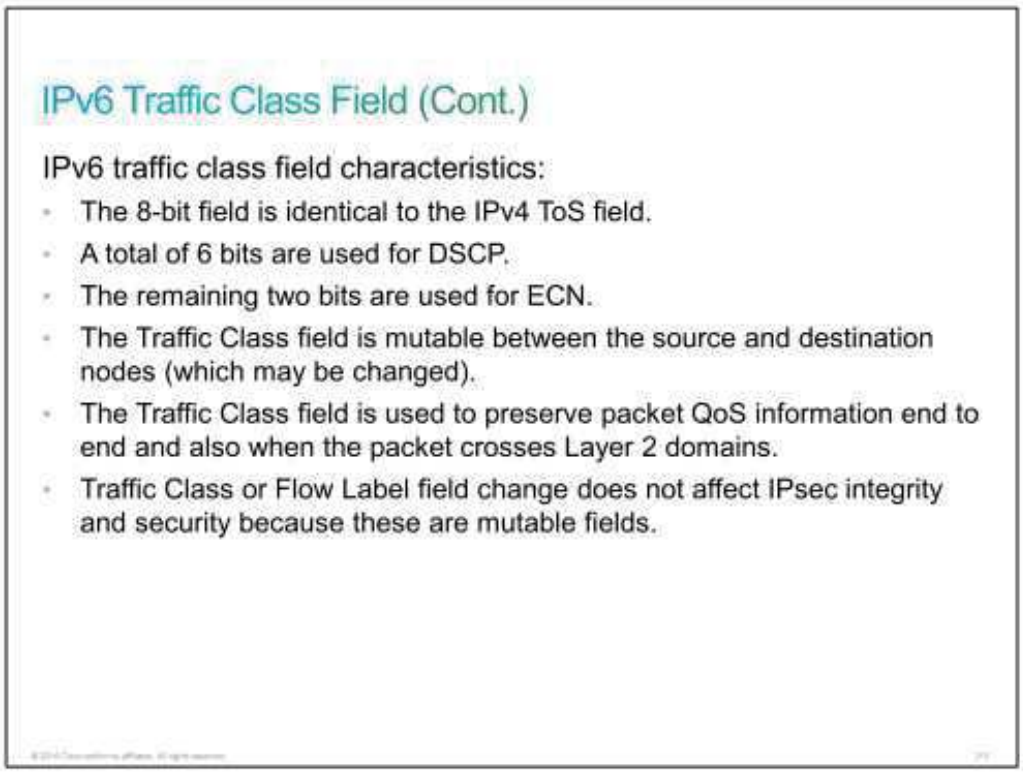
The IPv6 Traffic Class field is an 8-bit field that is identical to the ToS field in IPv4.




QoS field position has been moved toward the beginning to allow for easier hardware processing of packets. When receiving a packet, the network device can determine the priority of the packet very early in the process.



https://t.me/learningnets




Both the IPv6 Traffic Class and IPv4 ToS fields are used in the DiffServ architecture that is defined in RFC 2474, *Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers*. A total of 6 bits of the Traffic Class field are specified for use as the DSCP field, where each DSCP specifies a particular PHB that is applied to a packet at a network device.

The remaining 2 bits of the Traffic Class field are defined in RFC 3168, *The Addition of Explicit Congestion Notification (ECN) to IP*. ECN is a nonlossy way to indicate congestion on a link and to inform other systems to throttle traffic that is being sent to avoid packet drops due to congestion.

The Traffic Class field resides in the IPv6 packet header and *marks* the packets according to their priority. This means that the Traffic Class can be called a *Layer 3 marker*. Because the IPv6 header does not change in transit, this header is considered an end-to-end marker as well. Layer 2 markers, such as CoS for Ethernet networks, are valid only for a single Layer 2 domain.

If an IPv6 packet needs to cross multiple Layer 2 domains from its origin to its destination, the Traffic Class marker would be kept unchanged from end to end. New Layer 2 QoS markers would be imposed each time such a packet would be forwarded from one router to another over a Layer 2 network. However, the Traffic Class field can be changed on a Layer 3 device because this is a mutable field.

 ©2014 Cisco Systems, Inc.

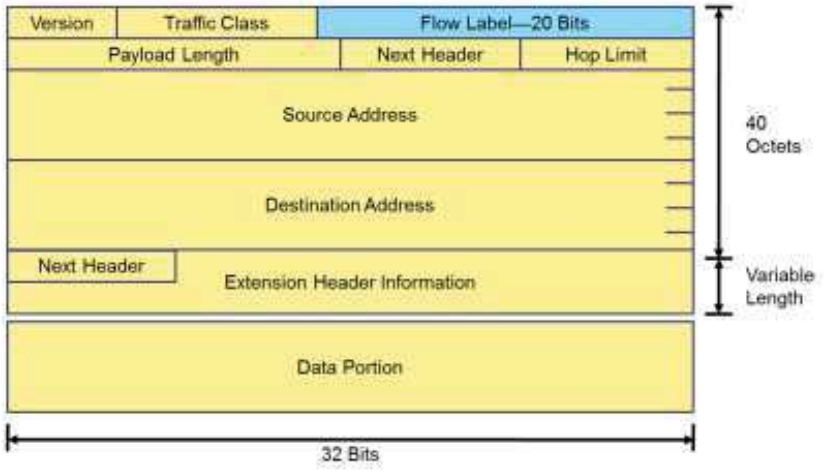
https://t.me/learningnets

# IPv6 Flow Label Field

The Flow Label field is a new 20-bit field that appears in the IPv6 basic header.

## IPv6 Flow Label Field

- Flow Label is a new IPv6 header field.




## IPv6 Flow Label Field (Cont.)

IPv6 flow label field characteristics:

- A new field is used to label packet flows.
- A flow can be used to request nondefault QoS.
- The Flow Label field is immutable between the source and destination nodes (which may not be changed, unlike the Traffic Class field).
- There are no existing implementations or standards defining the Flow Label field for QoS; they could be used to mark media streams.

The Flow Label field is used to label packets belonging to specific flows:

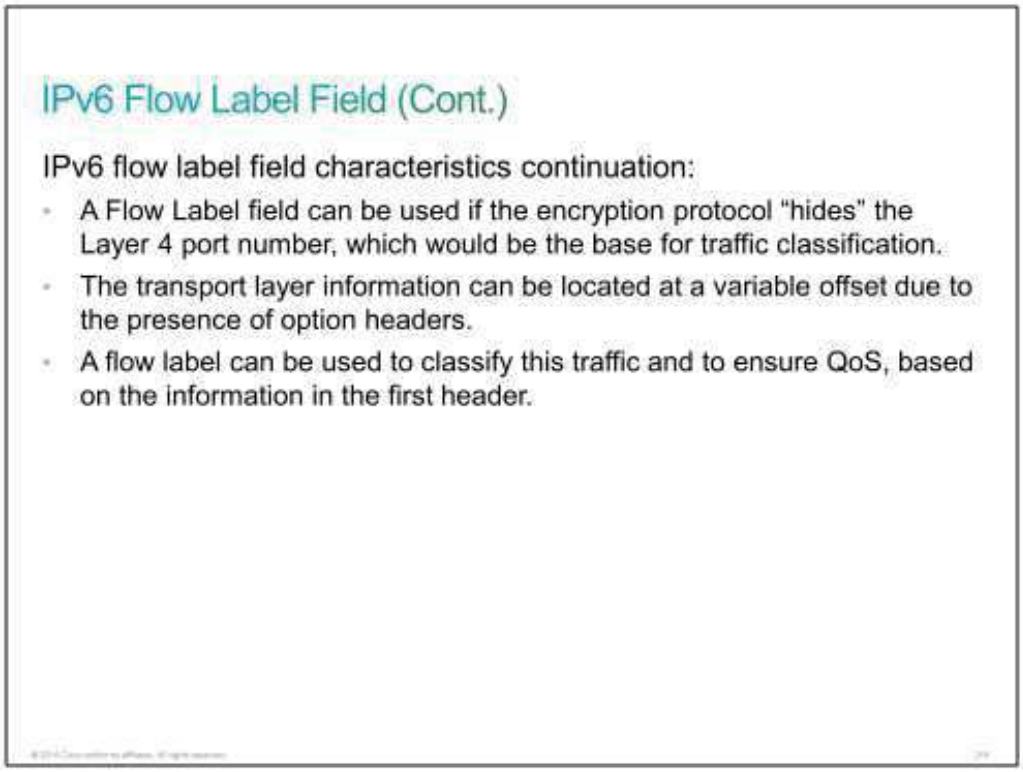


https://t.me/learningnets

- The flow can be uniquely identified based on source address, destination address, and flow label.
- The flow label can be used for special sender requests, such as nondefault QoS and real-time services.
- There can be multiple flows between a source and destination, as distinguished by separate nonzero flow labels.
- No implementation of the flow label currently exists, nor is its exact usage yet defined. There is a current IETF RFC that describes, at a high level, the basic requirements for the flow label (RFC 3697, *IPv6 Flow Label Specification*).
- The Flow Label field is immutable; its value must arrive at the destination unchanged.

The Flow Label field enables per-flow processing by routers in the path. This function provides differentiation of the traffic at the IP layer without having to open the transport layer header to identify the flow.

| | |
|---|---|
| **Note** | Consider a fragmented or encrypted packet. When a packet is fragmented, Layer 4 header information, such as TCP port number, is not carried in each fragment. In this case, for IPv4, QoS cannot be applied to each fragment when QoS classification is based on TCP port numbers. For IPv6 and the flow label, flows are classified only with information in the base header, which appears in each fragment. |
| **Note** | For IPsec-encrypted packets, Layer 4 information is encrypted and not available for QoS processing. The flow label and source or destination IP addresses are always visible in an encrypted packet, allowing QoS processing. |




The flow labels can be used to classify traffic if, for some reason, classification cannot be performed otherwise. As an example, packet encryption might obscure the transport layer headers that would otherwise be used for classification. Classification using flow labels could be useful.



https://t.me/learningnets

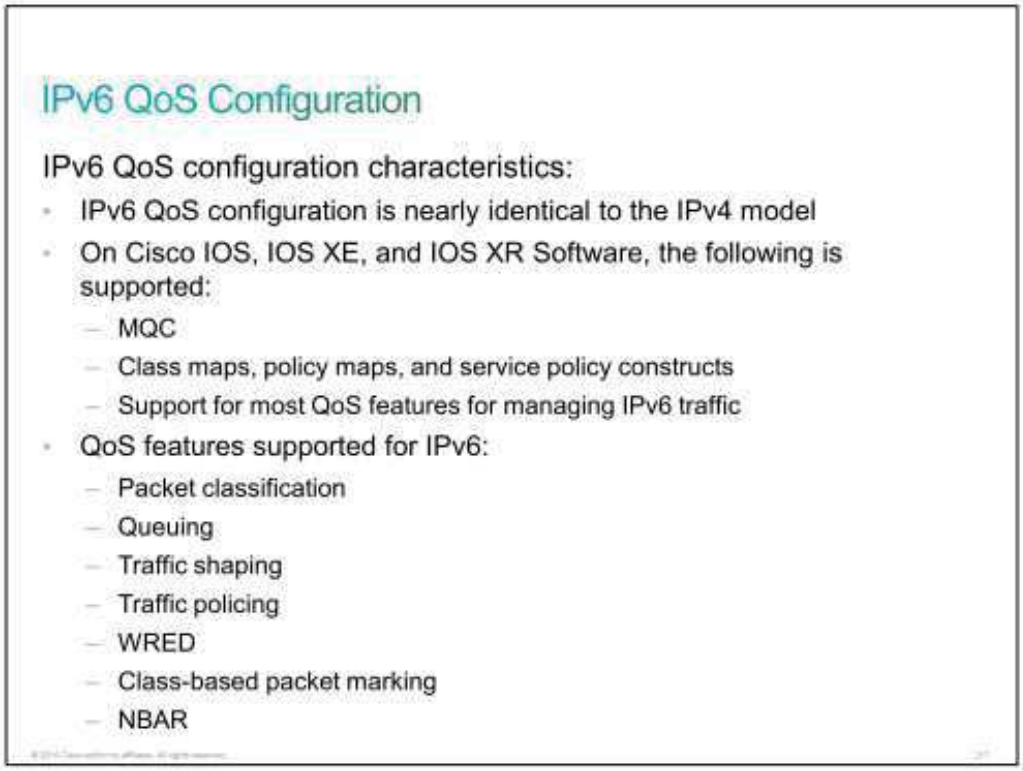
Next, to enable truly hardware-based QoS processing, the presence—or nonpresence—of option headers changes the position (offset) of a transport layer header. Using flow labels, these packets could be classified based on the information in the IPv6 (Layer 3) header.



https://t.me/learningnets

# IPv6 QoS Configuration

The IPv6 QoS configuration is described below:




Configuring IPv6 QoS in Cisco IOS, IOS XE, and IOS XR Software is nearly identical to configuring QoS for IPv4. These basic configuration elements are the same:

- MQC is supported
- Class maps, policy maps, and service policy commands are used

The MQC is a consistent, flexible way to configure QoS policies on Cisco routers. The three elements in the MQC that ease configuration tasks by enabling entire QoS policies to be applied to interfaces using a single command are as follows:

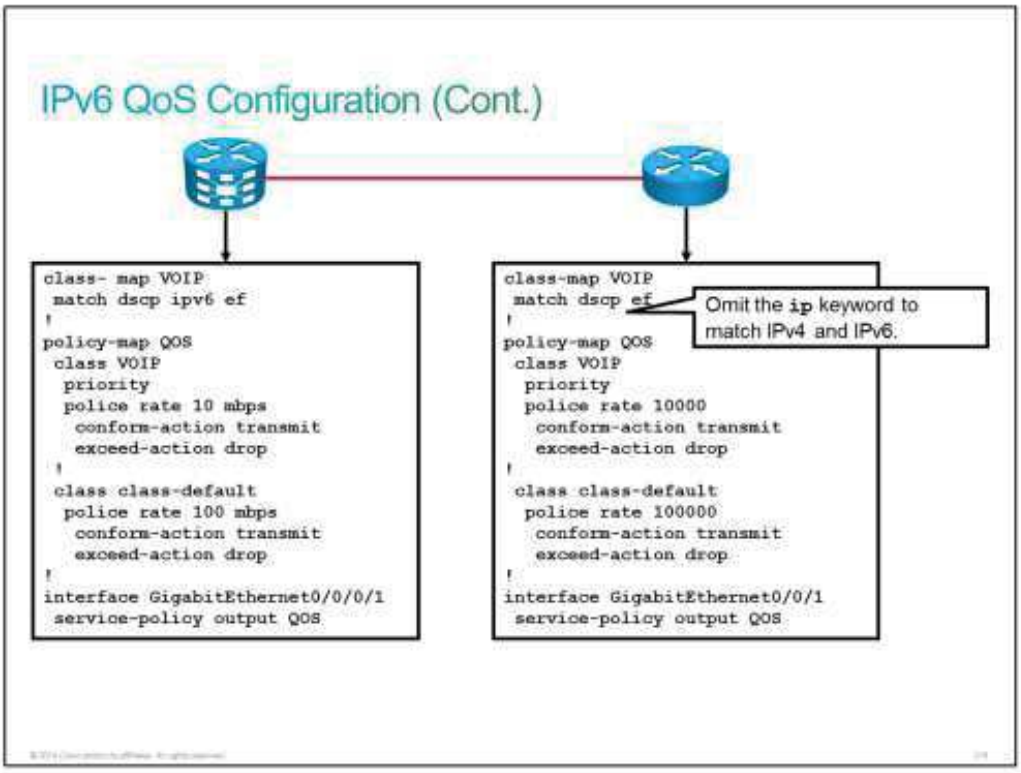
- **Class maps:** Class maps define which traffic to apply QoS policies to. Powerful classification commands are available to sort traffic that is based on Layer 2, 3, 4, or 7 criteria.
- **Policy maps:** Policy maps are used when QoS policies are applied to the traffic that is placed in the class maps that are previously defined. Packet marking (Layer 2 or Layer 3), policing and shaping, congestion avoidance, WRED, and various queuing methods are applied using policy maps. In addition, CBWFQ bandwidth guarantees can be applied, and priority queues can be established for real-time traffic.
- **Service policies:** Service policies apply policy maps to interfaces in a specific direction (inbound or outbound).

Cisco IOS Software supports many IPv4 QoS features for IPv6. Classification can be accomplished based on protocol (IPv4 or IPv6) or on protocol-independent values such as DSCP, CoS, or Layer 4 port.

 © 2014 Cisco Systems, Inc.

https://t.me/learningnets

Traffic policing and multiple shaping methods are, in addition to WRED, congestion avoidance. Most queuing methods are supported, including LLQ.




The figure shows a QoS sample configuration on the Cisco IOS XR and Cisco IOS and IOS XE Software platforms. The example enables LLQ for voice traffic and polices voice traffic to 10 Mb/s. Voice traffic is classified with the DSCP value set to EF. All other traffic is policed to 100 Mb/s. The QoS policy is applied to GigabitEthernet0/0/0/1 (that is, GigabitEthernet0/0/1 on the Cisco IOS Software router) in the outbound direction.

The difference when configuring QoS for IPv6 on the Cisco IOS and IOS XR Software platforms is that the Cisco IOS XR Software platform allows you to specify the IP version when matching DSPC or precedence values. Therefore, you can match only DSCP or precedence values that are inside IPv4 or IPv6 packets. On Cisco IOS Software, you can match either IPv4 values only (by providing the **ip** keyword in front of the **dscp** or **precedence** keyword) or both IPv4 and IPv6 values by omitting the **ip** keyword.

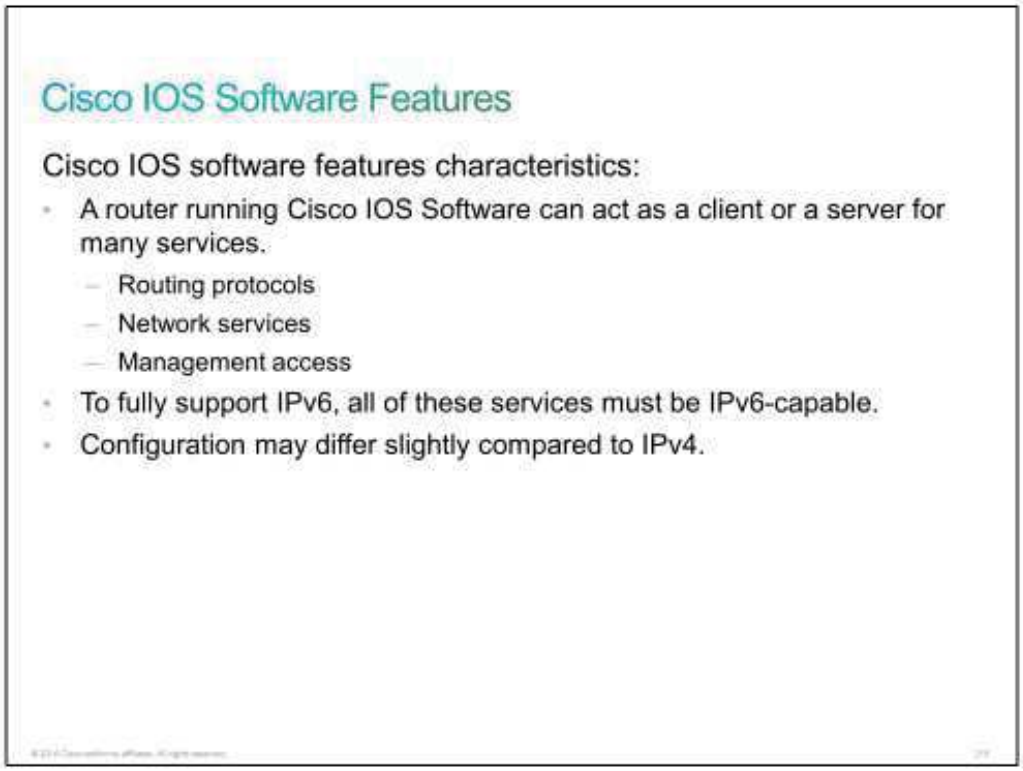
| | |
|---|---|
| **Note** | For a complete command reference, refer to the Cisco IOS, IOS XE, and IOS XR Software Command Reference guides on www.cisco.com. |



https://t.me/learningnets

# Cisco IOS Software Features

The Cisco router is not just forwarding packets. It also runs routing protocols that may need to communicate over IPv6.




In addition, a router also can offer various network services, can act as a troubleshooting platform, and must support various protocols for administrative access.

In an IPv6-only network, all of these features must fully support IPv6 as a transport mechanism.

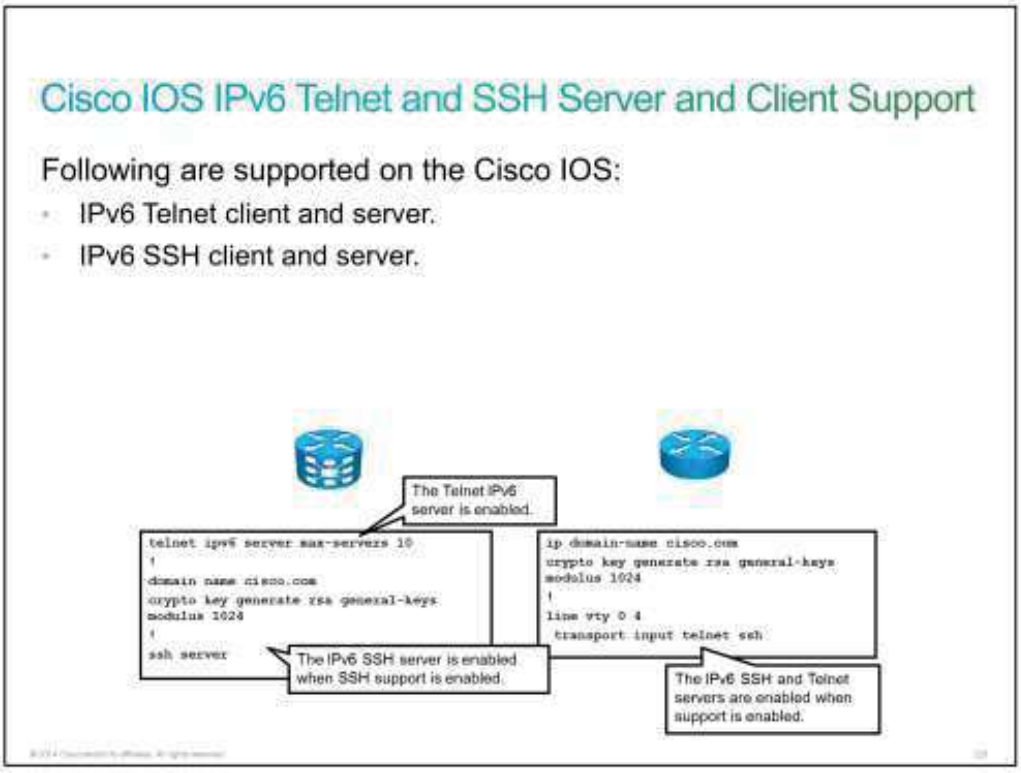
In most cases, IPv6 configuration commands do not differ significantly from their IPv4 counterparts. Sometimes, the commands are the same. In some cases, however, configuring IPv6 and configuring IPv4 are substantially different.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# Cisco IOS IPv6 Telnet and SSH Server and Client Support

The IPv6 Telnet and SSH server and client support in Cisco IOS Software is described below:




Telnet is a primary method of doing remote management. Unfortunately, Telnet is insecure because Telnet data goes over the network in cleartext.

The Telnet client and server support IPv6 connections. You can connect directly to a router using an IPv6 Telnet client. You can also initiate an IPv6 Telnet connection from a router.

SSH is a popular replacement for Telnet because it provides security features that are not available on Telnet. In a plain Telnet user session, the authentication and the session are transported in cleartext over the network. Anyone in the path between the user and the router can intercept the session information.

SSH protects the interactive session through encryption and also can be used to provide stronger authentication mechanisms and other features. SSH is available for both IPv4 and IPv6 when running an IPsec-capable version of Cisco IOS Software.

The figure shows how to enable the Cisco IOS and IOS XR Software routers for SSH and Telnet for IPv6.

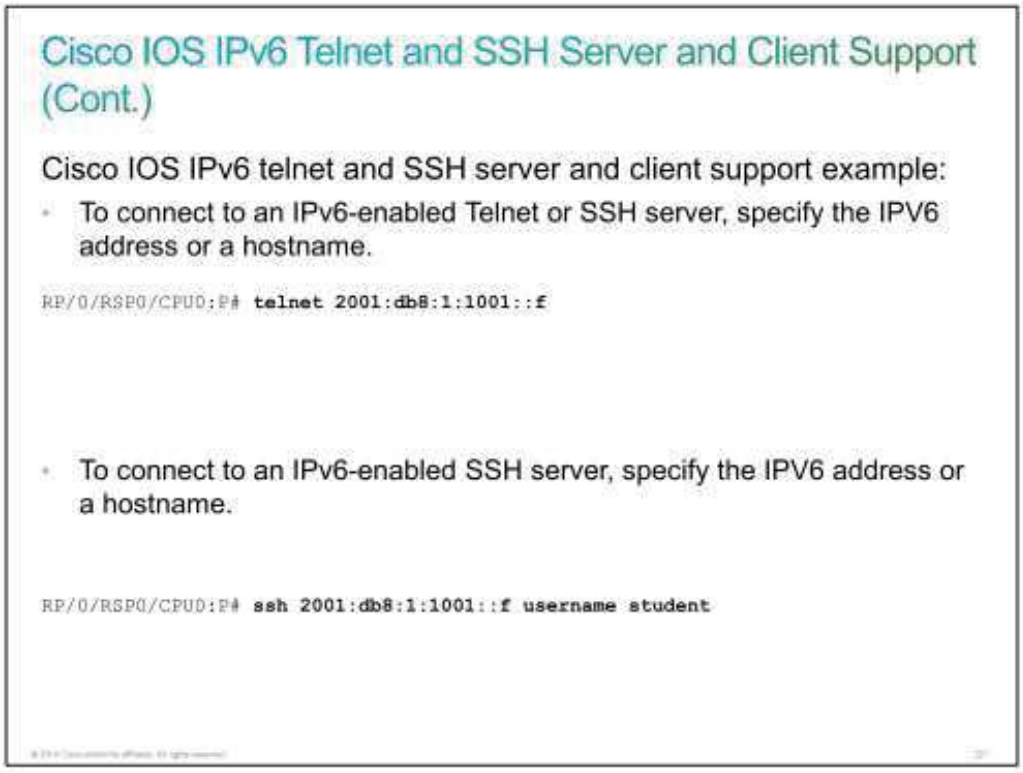
On the Cisco IOS XR Software router, Telnet for IPv6 is enabled using the **telnet ipv6 server max-servers** command, followed by a number of maximum concurrent sessions to the router. SSH for IPv6 is enabled the same way as for IPv4. First configure the hostname and the domain name on the router. Then create an RSA key that will be used for signing and encryptions using the **crypto key generate rsa general-keys modulus** command. Then enable the SSH server using the **ssh server** command.

On the Cisco IOS and IOS XE Software routers, SSH and Telnet for IPv6 are enabled the same way as for IPv4.



https://t.me/learningnets

| Note | For a complete command reference, refer to the Cisco IOS, IOS XE, and IOS XR Software Command Reference guides on www.cisco.com. |
|---|---|



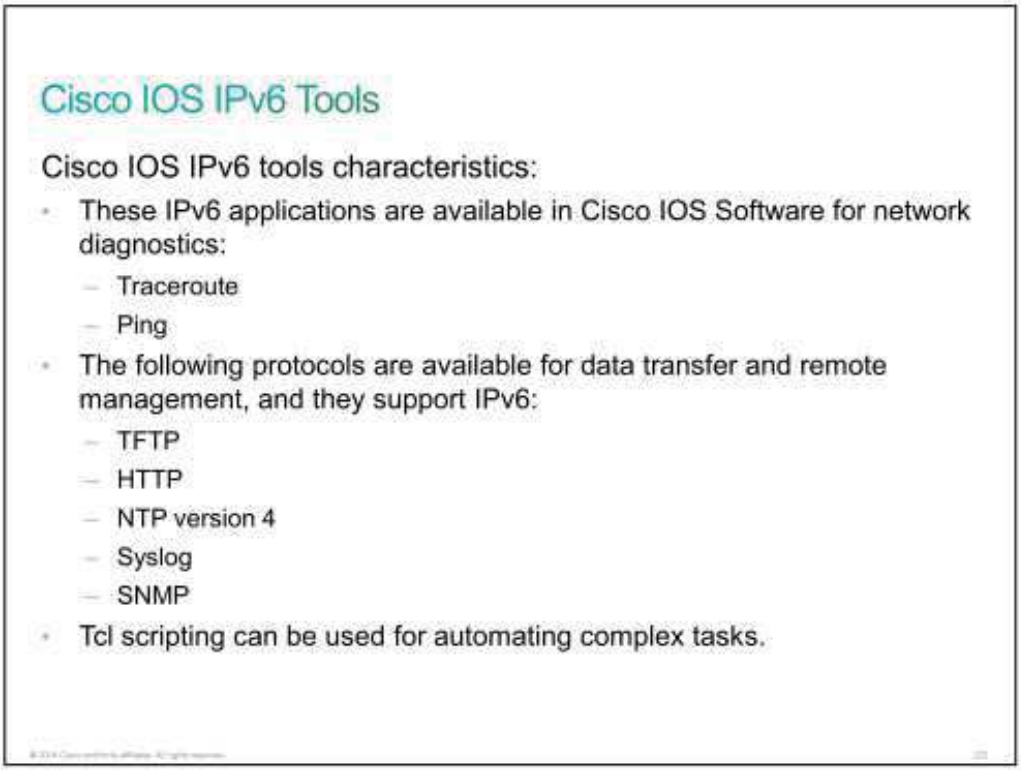

To use the Telnet or SSH client on the Cisco router over IPv6 and to connect to an IPv6-enabled Telnet or SSH router, provide an IPv6 address or a hostname that resolves into the IPv6 address as an argument to **telnet** and **ssh** commands.



https://t.me/learningnets

# Cisco IOS IPv6 Tools

Other IPv6 tools in Cisco IOS are described below:




These IPv6-enabled applications are available on Cisco IOS Software routers:

- **traceroute** *destination*: This command accepts a destination IPv6 address or hostname as an argument and generates IPv6 traffic to report each IPv6 hop that is used to reach a destination address.
- **ping** *destination*: This command is used to send an ICMPv6 echo request to a destination. The ICMPv6 echo reply is reported on the console. Extended ping is also supported.

These applications can take an IPv6 address or a hostname as an argument that will resolve to an IPv6 address. (Currently, DNS AAAA records are supported, but the A6 record is not.)

The following protocols are used on Cisco routers for network management:

- **TFTP:** TFTP is used for sending or receiving files, most often configuration settings or software binaries. It transparently supports IPv6 addresses. A router can act as a client or as a server.
- **HTTP:** You can use a web browser to connect to a router over IPv6. The configuration page that is shown in the browser window will be the same as that for web-based connections that are made over IPv4.
- **NTP:** NTP is used for time synchronization. The correct time can be very important when comparing logs from different devices and when certificates are used for authentication. NTPv4 adds support for IPv6.
- **SNMP:** SNMP is used for remote management and reporting. It is often used for status monitoring. SNMP fully supports IPv6 as a transport protocol.



https://t.me/learningnets

- **Syslog:** Syslog is used for sending log messages to a remote host over UDP. Syslog supports IPv6 as a transport protocol.

Tcl is a scripting language that is implemented in Cisco IOS Software and can be used for automating tasks. It can integrate with ESM, Cisco IOS Embedded Event Manager, and interactive voice response IVR.

All of these applications and protocols support IPv6 and take IPv6 addresses as arguments when configuring them.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# Cisco Discovery Protocol Support for IPv6

Cisco Discovery Protocol is commonly used to discover the protocol addresses of neighboring devices and the platform of those devices. Cisco Discovery Protocol is protocol independent and supports IPv6 address and address type information.

## Cisco Discovery Protocol Support for IPv6

Cisco Discovery Protocol characteristics:

- Cisco Discovery Protocol:
  - Used to discover protocol addresses and platform information of neighboring devices.
  - Runs on all Cisco devices (that is, routers, switches, and so on).
- Cisco Discovery Protocol IPv6 support:
  - Adds IPv6 address and address type information.

## Cisco Discovery Protocol Support for IPv6 (Cont.)

- Displays detailed information about neighbors.

```
RP/0/RSP0/CPU0:P# show cdp neighbors detail
<… output omitted …>
-------------------------
Device ID: PE8.lab.com
SysName :
Entry address(es):
IPv4 address: 192.168.82.80
IPv6 address: 2001:db8:192:168:82::80
IPv6 address: fe80::207:7dff:fe33:2783
Platform: cisco ASR1001, Capabilities: Router IGMP
Interface: GigabitEthernet0/0/0/9
Port ID (outgoing port): GigabitEthernet0/0/3
Holdtime : 176 sec
```



https://t.me/learningnets

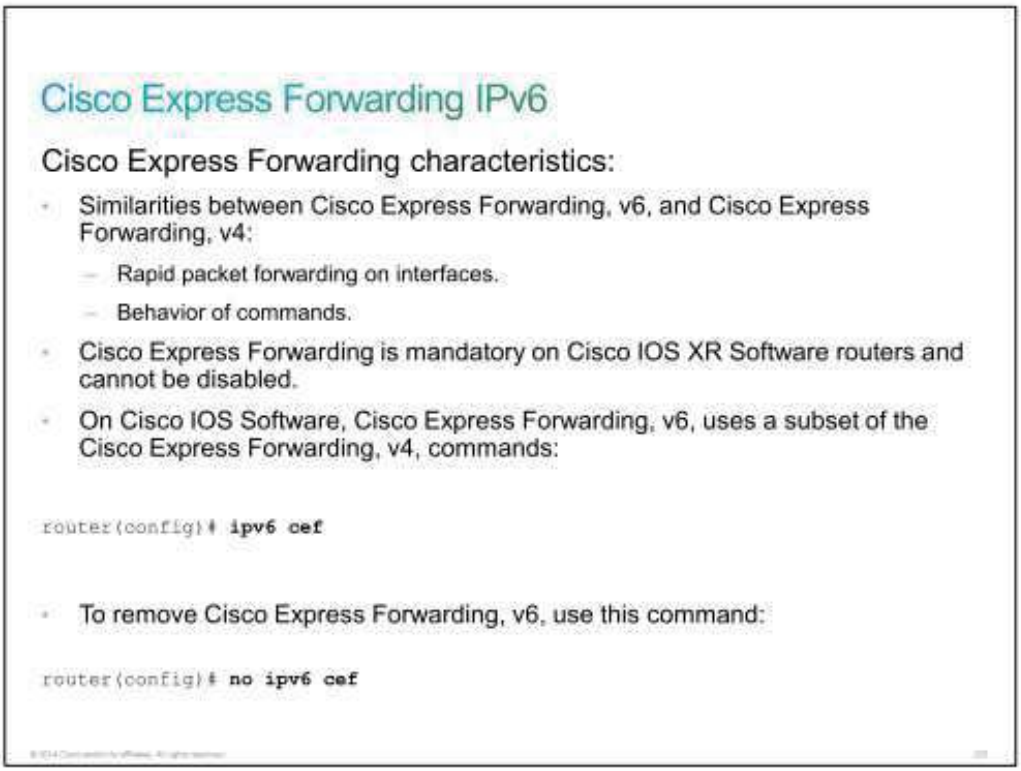
This example demonstrates the output of the **show cdp neighbors detail** command. Note that the neighbor device, PE8, has two IPv6 addresses on its interface: one link-local and one configured.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# Cisco Express Forwarding for IPv6

Cisco Express Forwarding is a Layer 3 switching technology that is designed for routers. It uses a method that optimizes route lookups to achieve very fast traffic forwarding. Cisco Express Forwarding uses two tables to store the information that is needed for routing: the FIB and the adjacency table.




The behavior of Cisco Express Forwarding version 6 is the same as that of Cisco Express Forwarding version 4. There are new configuration commands for Cisco Express Forwarding version 6 and common commands for both Cisco Express Forwarding, version 6, and Cisco Express Forwarding, version 4.

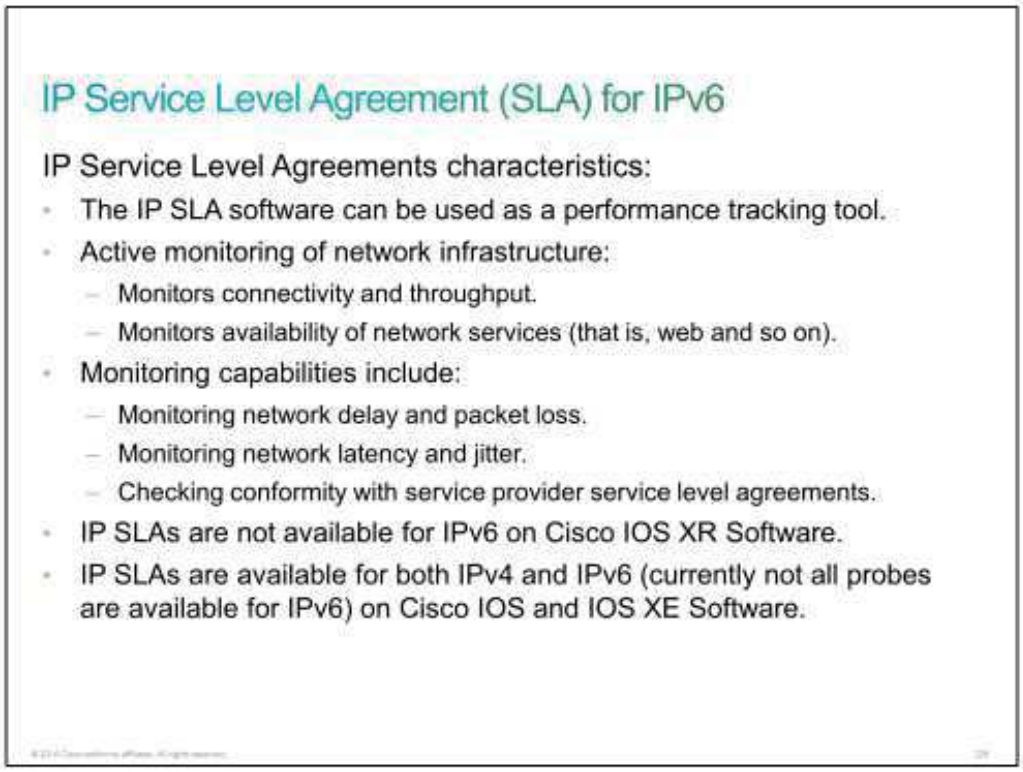
On the Cisco IOS XR Software platforms, Cisco Express Forwarding is mandatory and cannot be disabled. On Cisco IOS and IOS XE Software platforms, Cisco Express Forwarding IPv6 uses subset of Cisco Express Forwarding IPv4 commands. For example, the **ipv6 cef** command enables the central Cisco Express Forwarding, version 6, mode. However, IPv4 Cisco Express Forwarding must first be enabled using the **ip cef** command.



https://t.me/learningnets

# IP Service Level Agreement (SLA) for IPv6

Cisco IOS IP SLA is a part of Cisco routers, which allows customers to analyze IP service levels for IP applications and services, to increase productivity, and to reduce the frequency of network outages.




Cisco IOS IP SLA uses active traffic monitoring—the generation of traffic in a continuous, reliable, and predictable manner—to measure network performance. Using Cisco IOS IP SLAs, customers can verify service levels verify internal and outsourced SLAs, and understand network performance.

Cisco IOS IP SLAs can perform network assessments, verify QoS, ease the deployment of new services, and assist administrators with network troubleshooting.

IP SLA for IPv6 is not currently available on Cisco IOS XR Software platforms. It is available (with some exceptions) on Cisco IOS and IOS XE Software platforms.

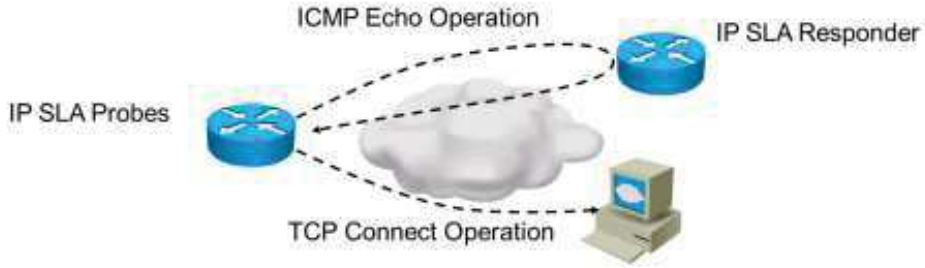
 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

## IP Service Level Agreement (SLA) for IPv6 (Cont.)

IP SLA key components:

- The IP SLA architecture consists of an IP SLA source and an IP SLA target.
- A probe is configured on the source, checking connectivity from the source to the target.
- The target device can be a router with an IP SLA responder or an IPv6 endpoint.




Cisco IOS IP SLA sends data across the network to measure performance between multiple network locations or across multiple network paths. It simulates network data and IP services and collects network performance information in real time. The information that is collected includes data about response time, one-way latency, jitter (interpacket delay variance), packet loss, voice quality scoring, network resource availability, application performance, and server response time.

Cisco IOS IP SLA performs active monitoring by generating and analyzing traffic to measure performance either between Cisco IOS devices or from a Cisco IOS device to a remote IP device such as a network application server. Measurement statistics that are provided by the various Cisco IOS IP SLA operations can be used for troubleshooting, for problem analysis, and for designing network topologies.

Cisco IOS IP SLA starts when the Cisco IOS IP SLA device sends a generated packet to the destination device. After the destination device receives the packet, and depending on the type of Cisco IOS IP SLA operation, the device will respond with time-stamp information for the source to make the calculation on performance metrics. A Cisco IOS IP SLA operation performs a network measurement from the source device to a destination in the network using a specific protocol such as UDP.

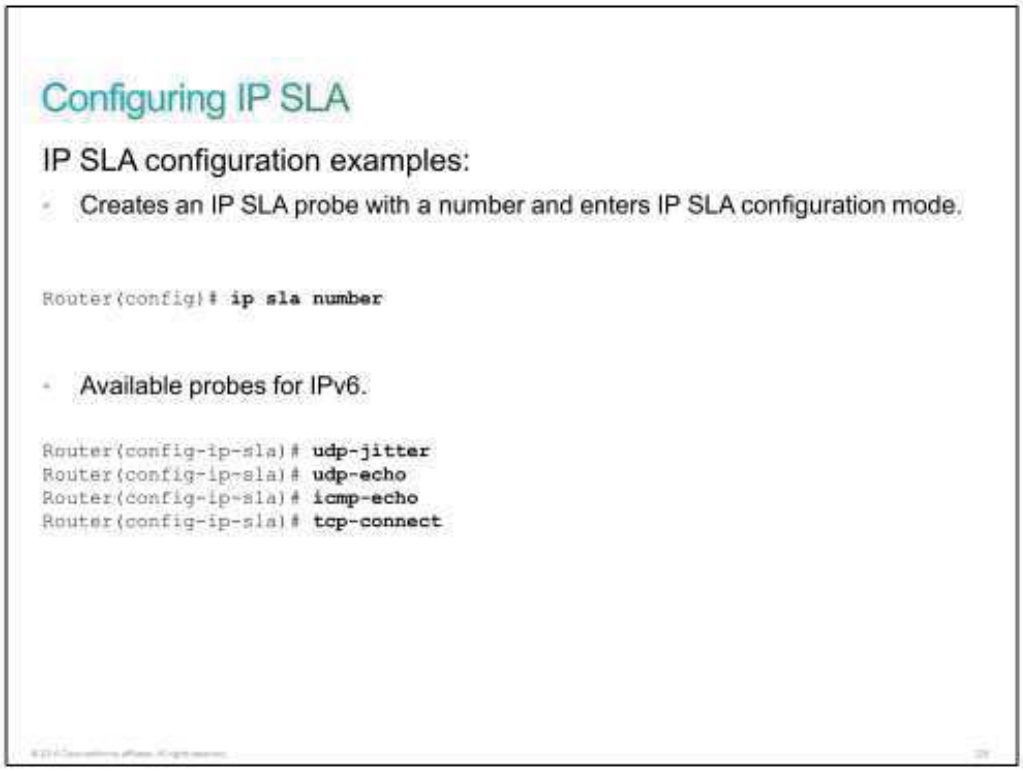
| **Note** | To use IP SLA between a customer and a service provider, the customer and the service provider have to have an agreement to configure IP SLA on both ends. |
|---|---|

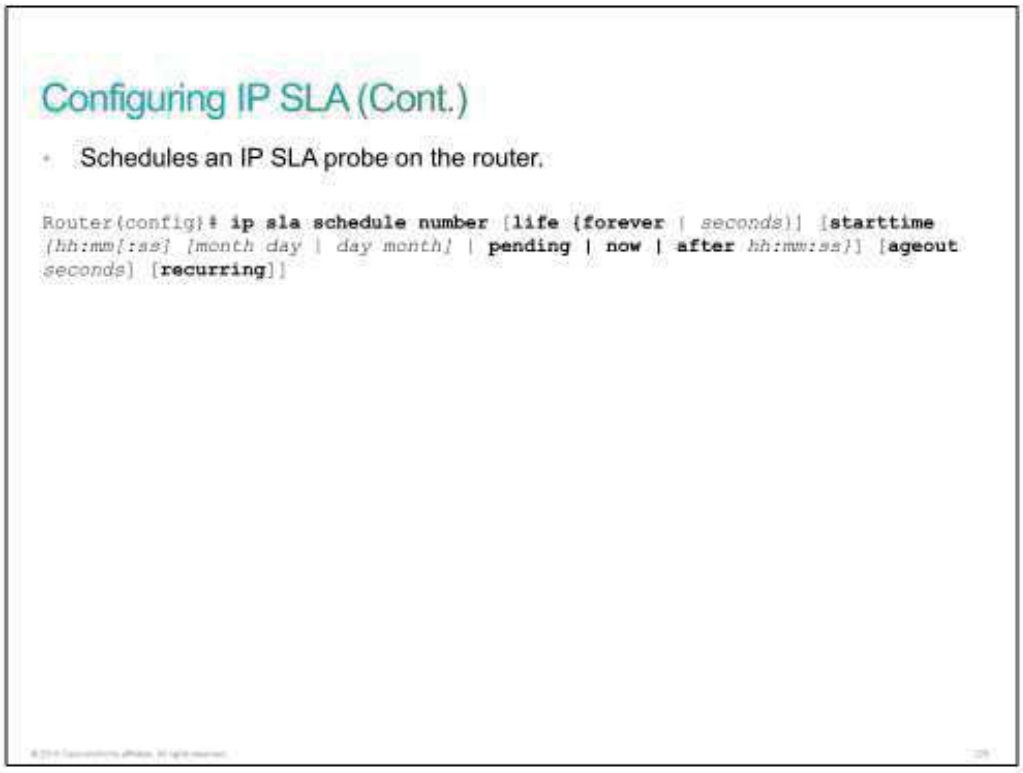


https://t.me/learningnets

# Configuring IP SLA

IP SLA is configured by creating an IP SLA probe on the source device and by scheduling this probe to run at the desired time.




 ©2014 Cisco Systems, Inc.

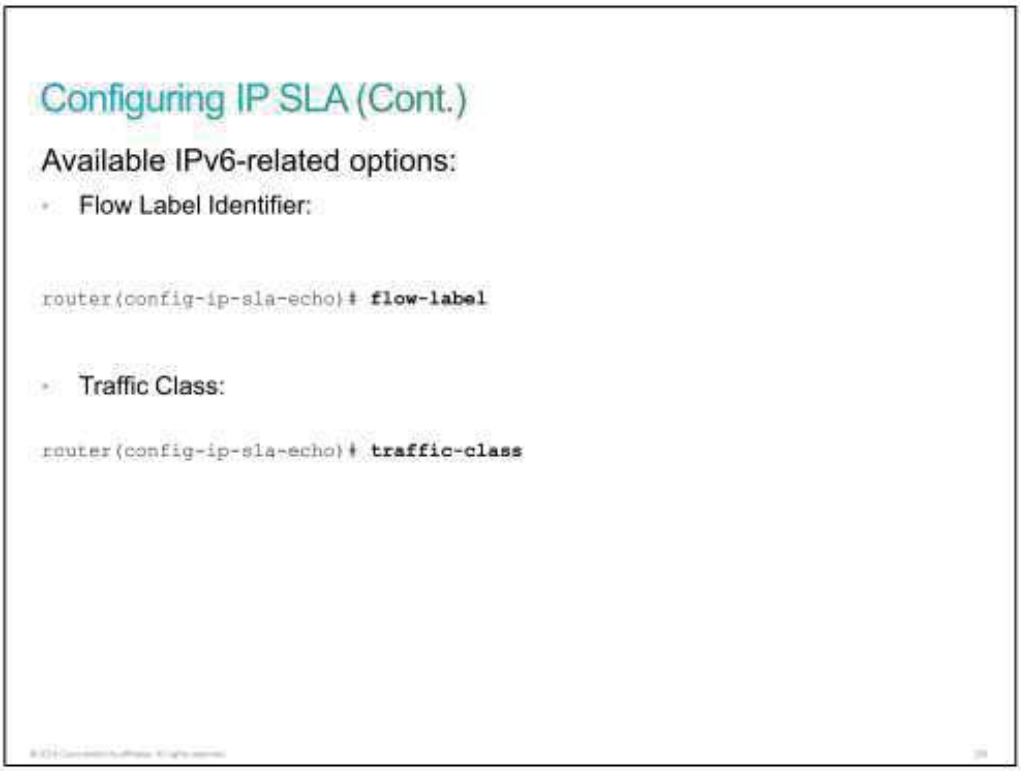
https://t.me/learningnets

The commands for the currently available probes of IP SLA that can be used for IPv6 are **udp-jitter, udp-echo, icmp-echo**, and **tcp-connect**. See the table for a description of functionality of these probes:

| IP SLA Operation | Measurements | Key Monitoring Application |
|---|---|---|
| UDP Jitter | Measures round-trip delay, one-way delay, one-way jitter, one-way packet loss, and connectivity testing of networks that carry UDP traffic, such as voice<br>Note: One-way delay requires time synchronization between the source and target routers. | Voice and data network performance<br>Note: These are the most commonly used Cisco IOS IP SLA operations. |
| | | General IP performance |
| UDP Echo | Measures round-trip delay of UDP traffic | Server and IP application performance |
| | | Connectivity testing |
| ICMP Echo | Measures round-trip delay of the full path | IP performance |
| | | Connectivity measurement |
| TCP Connect | Measures the time taken to connect to a target device with TCP | Server and application performance |

To schedule an IP SLA probe, use the **ip sla schedule** command.

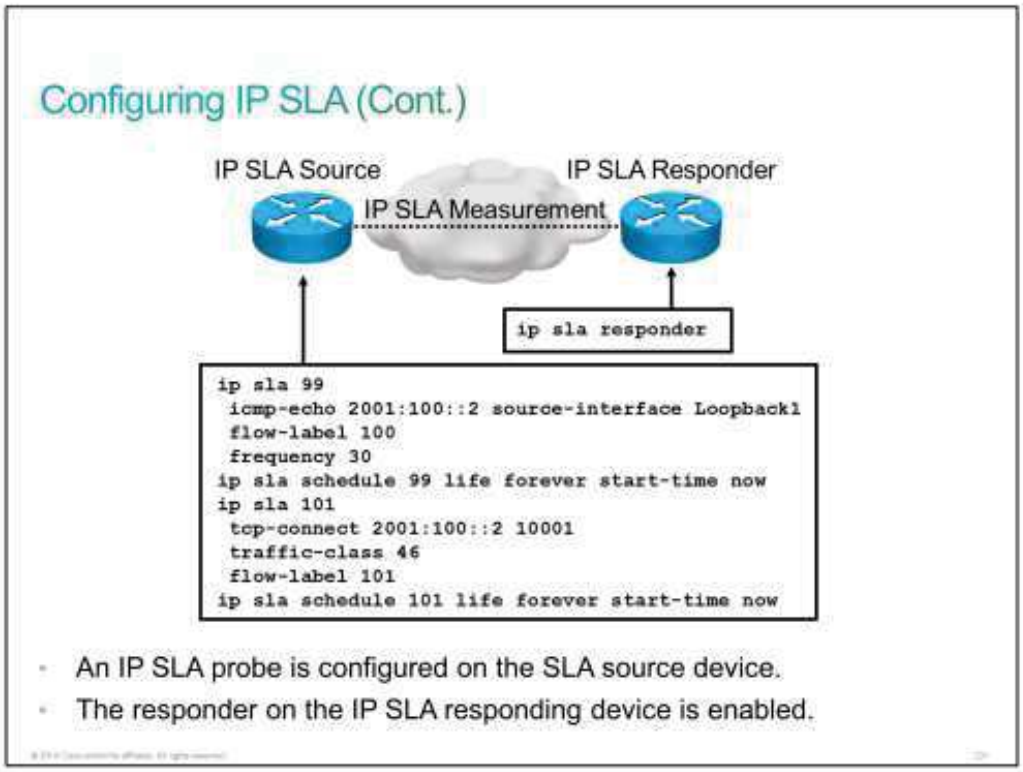
The probe will start functioning after the IP SLA responder process has been started on the destination router.






https://t.me/learningnets

When configuring probe options, you can set an IPv6 flow label or an IPv6 traffic class for the packets of the probe. Flow labels allow for easy identification of IPv6 packets, and traffic class allows the probe packets to be a high priority, not to be dropped during congestion.




The figure shows a sample configuration of IP SLA if configured using the command line. On the IP SLA source device, the probe must be configured, while on the SLA responder device, only the responder must be enabled.

When IP SLA sends the packets, it uses a time stamp to determine when the packet was sent and how much time the packet needed to travel across the networks. Based on this information, SLA statistics can be calculated.

The example shows ICMP echo and TCP connection probes that are configured on the IP SLA source device. The target device in this case can be an SLA responder, a target end device responding to ICMP echo packets, or TCP-initiated connections.

| | |
|---|---|
| **Note** | For a complete command reference, refer to the Cisco IOS and IOS XE Software Command Reference guides on www.cisco.com. |



https://t.me/learningnets

# Summary

This topic summarizes the key points that were discussed in this lesson.

## Summary

- Multicast is used in the infrastructure layer of the Cisco IP NGN.
- IPv6 multicast addresses start with "FF".
- IPv6 multicast address scope defines the reach of the multicast address group.
- Solicited node multicast address is used when doing neighbor discovery.
- Global scope represents the entire Internet.
- Multicast in IPv6 is conceptually similar to multicast in IPv4. The biggest difference is in domain control.
- PIM is protocol independent: MRIB is built on underlying unicast routing table.
- The RP address can be embedded into a multicast group address.

## Summary (Cont.)

- To enable multicast routing use multicast-routing IOS XR command or ipv6 multicast-routing IOS/IOS XE command.
- MLD handles join and leave processes on the access segment between the listener and the first multicast router.
- MLD v1 messages are Query, Report and Done.
- MLDv1 Query: "Which multicast addresses have listeners on the link?"
- MLDv1 Report: Nodes reply with the list of multicast groups they are receiving.
- MLDv1 Done: A done message is a node leaving a group.
- Multicast-address-specific queries have their Maximum Response Delay set to Last Listener Query Interval.
- MLDv2 defines new messages.



https://t.me/learningnets

## Summary (Cont.)

- Access groups enable access control for multicast group receivers on a MLD router.
- MLD join-group and static-group allow routers to join multicast groups.
- MLD is used with IPv6 multicast traffic to handle join and leave processes on the access segment between the listener and the first multicast router.
- Snooping of MLD packets is used to determine and store the information that switch ports receive data for multicast groups.
- DNS and DHCPv6 are used in the services layer of the Cisco IP NGN.
- DNS replaces A record with AAAA record for IPv6 addresses.
- Dynamic DNS works on the same principle as in IPv4.
- Before using DHCPv6, the client will check for the presence of any routers.

## Summary (Cont.)

- DHCPv6 cannot send gateway information to the client.
- DHCPv6-Lite is used to send DNS information to the client, without configuring the client address.
- DHCPv6 can be used to delegate prefixes to client routers instead of single addresses.
- DHCPv6 uses similar commands as IPv4 DHCP.
- QoS is used on the IP infrastructure layer of the Cisco IP NGN.
- IPv6 keeps ToS fields under new name Traffic class. It has an additional flow label field.
- Traffic class field is just renamed ToS field.
- Flow label field is new addition to IPv6 and is used to distinguish flows at layer 3.


© 2014 Cisco Systems, Inc.

https://t.me/learningnets

## Summary (Cont.)

- QoS can be configured to be protocol agnostic.
- To simplify transition to IPv6, systems should be dual stacked to support IPv6 in conjunction with IPv4.
- Cisco IOS SSH and telnet support IPv6 seamlessly.
- Most other tools in Cisco IOS also support IPv6.
- Cisco Discovery Protocol supports IPv6.
- To take the full advantage of IPv6 forwarding you must enable Cisco Express Forwarding for IPv6.
- IP SLA supports IPv6, but not all features may be available.
- The commands for the currently available probes of IP SLA that can be used for IPv6 are **udp-jitter, udp-echo, icmp-echo**, and **tcp-connect**.



https://t.me/learningnets

https://t.me/learningnets

# Lesson 2: Defining IPv6 Transition Mechanisms

## Overview

Due to near depletion of public IPv4 addresses, service providers have to take measures in their environments to fight against it. The ultimate solution is to offer IPv6 connectivity to the customer. When enabling IPv6 support in service provider environments, you are required to make choices about the method or methods that are used to support the integration. Numerous approaches are available, and you have to select the appropriate one for the transition scenario.

This lesson first describes dual stack. The lesson then continues with how dual-stack implementations can be improved using CGN and how NAT64 can be used for IPv6-only hosts. The lesson concludes with a description and configuration of various IPv6 tunneling technologies that enable connectivity between standalone IPv6 islands with IPv6 Internet-over-IPv4 networks.

Upon completing this lesson, you will be able to:

- Describe dual-stack implementation, CGN, and NAT64
- Describe dual-stack operations
- Describe dual-stack considerations
- Describe dual stack with carrier-grade NAT operations
- Describe NAT444 operations
- Describe carrier-grade NAT support on Cisco routers
- Describe NAT64
- Describe DNS64 operations
- Describe stateless NAT64 operations
- Describe stateful NAT64 operations
- Compare stateless and stateful NAT64



https://t.me/learningnets

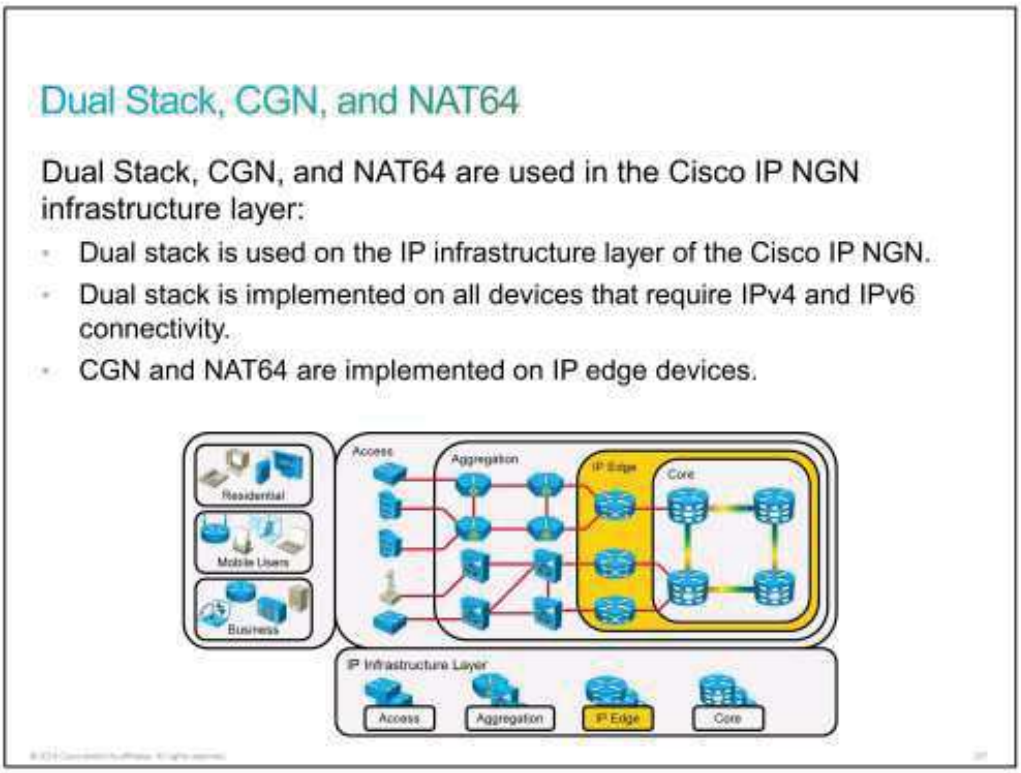
- Describe the stateful NAT64 configuration on the ASR 1000
- Describe the static, stateful NAT64 configuration on the ASR 1000
- Describe IPv6 tunneling mechanisms
- Describe IPv6 manual tunnels
- Describe GRE manual tunnels
- Describe 6in4 manual tunnels
- Describe 6in4 manual tunnel configuration
- Describe 6to4 automatic tunnels
- Describe 6to4 considerations
- Describe 6to4 configuration
- Describe 6RD automatic tunnels
- Describe 6RD considerations
- Describe 6RD configuration

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# Dual Stack, CGN, and NAT64

Dual stack is used on the IP infrastructure layer of the Cisco IP Next-Generation Network. Dual stack is required on all devices that require native access to IPv4 and IPv6 resources. If dual stack is not implemented throughout the network, NAT64 can be used to enable users to access IPv4 content from the IPv6 network. If there is public IPv4 address depletion on the service provider side, the service provider can use the CGN solution to assign private IPv4 addresses to customers. These technologies are implemented on IP edge devices in the P-network.



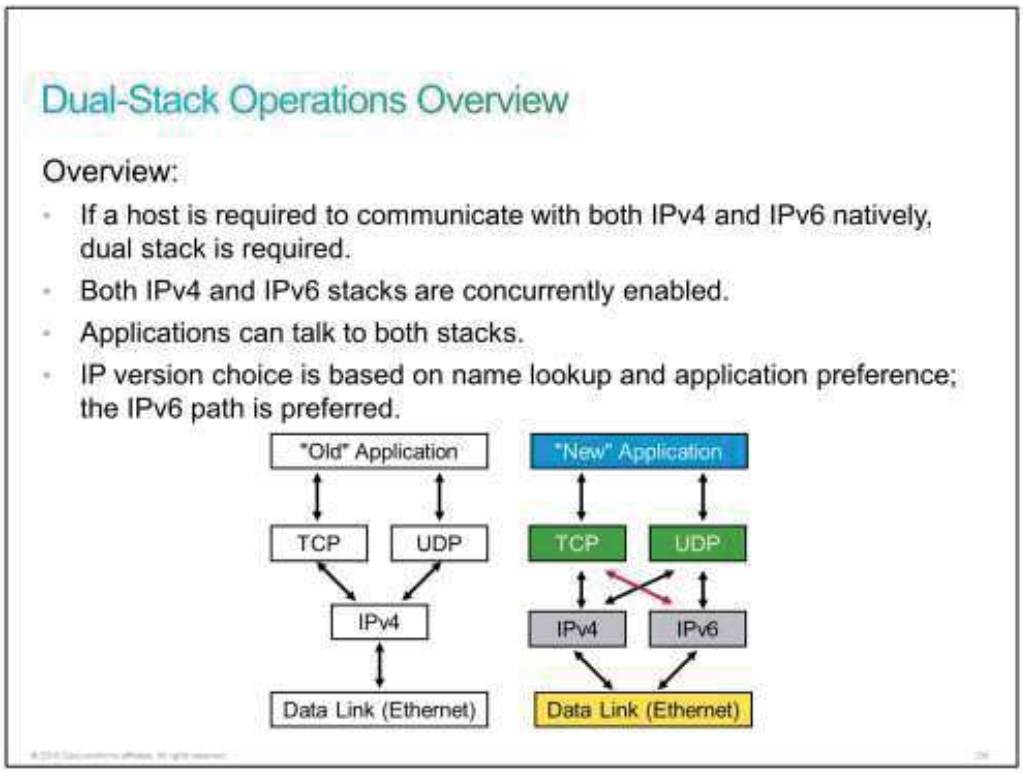

The figure shows where dual stack, CGN, and NAT64 are used in the Cisco IP NGN infrastructure layer.



https://t.me/learningnets

# Dual-Stack Operations Overview

The preferred strategy for transitioning to IPv6 is the dual-stack approach, where nodes have both IPv4- and IPv6-enabled stacks; some applications must be modified to use IPv6. "Old" IPv4-only applications continue to work as before. "New" and modified applications take advantage of both IP layers, usually preferring to use the IPv6 protocol for communication.




Past experience in porting IPv4 applications to IPv6 shows that, for most applications, the transition requires a minimal change in some localized places inside the source code.

The dual-stack technique is well known and has been applied in the past for other protocol transitions. It enables one-by-one application upgrades to IPv6.



https://t.me/learningnets

## Dual-Stack Operations Overview (Cont.)

A dual-stack operation is an application that is not aware of IPv6, or if the destination is only IPv4-enabled:

- It asks the DNS for an IPv4 address.
- It connects to the IPv4 address.
- Legacy applications exist; new applications should be designed as protocol-independent.




The figure shows the typical situation without IPv6.

An application that is not aware of IPv6, or is forcing the use of IPv4, requests only the IPv4 address of the destination hostname (http://www.example.com) from the DNS. The DNS request for an IPv4 address is a request for an A record. The DNS server sends back only the IPv4 address of the host. The application then requests the source host to connect to the destination host using IPv4.

## Dual-Stack Operations Overview (Cont.)

An application or destination that is only IPv6-enabled has to use IPv6 because it is the only stack:

- It asks the DNS for an IPv6 address.
- It connects to the IPv6 address.




The figure shows the typical situation for IPv6-only or IPv6-preferred services.

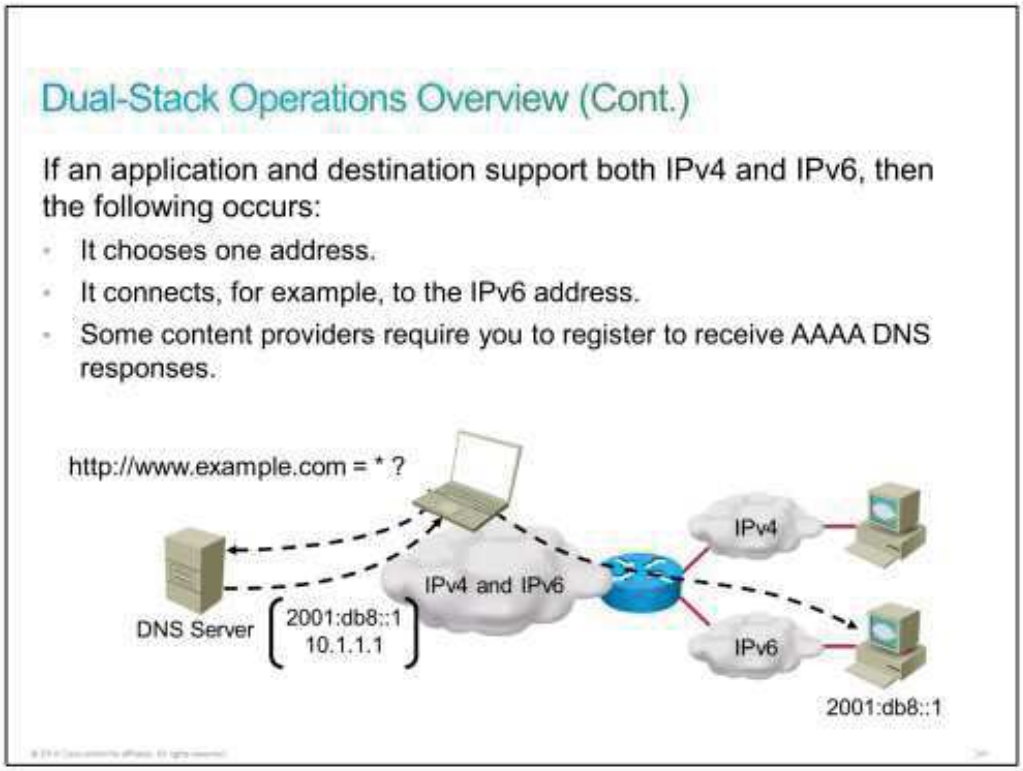


https://t.me/learningnets

The application requests only the IPv6 address of the destination hostname (http://www.example.com) from the DNS when an application is one of the following:

- The application is only IPv6-enabled.
- IPv6 is the only stack available on the host.
- The application is forcing the use of IPv6.

The DNS request for an IPv6 address is a request for an AAAA record. The DNS server sends back only the IPv6 address of the host. The application then requests the source host to connect to the destination host using IPv6.




The figure shows the typical situation for dual-stack hosts with applications supporting both protocols.

An IPv6-enabled and IPv4-enabled application sends requests to the DNS for all types of addresses for the destination hostname (http://www.example.com). The DNS server sends all available host addresses (IPv4 and IPv6). The application chooses one or the other (the typical default is IPv6) and then requests the source host to connect to the destination host using IPv6.

| | |
|---|---|
| **Note** | RFC 3484, *Default Address Selection for Internet Protocol version 6 (IPv6)*, describes how nodes choose a destination address from a set of IPv4, and possibly IPv6, destination addresses and how nodes choose the source address to use for a connection. The rules in this RFC prefer IPv6 over IPv4, when both are available, and suggest that nodes should be able to change this default behavior. |



https://t.me/learningnets

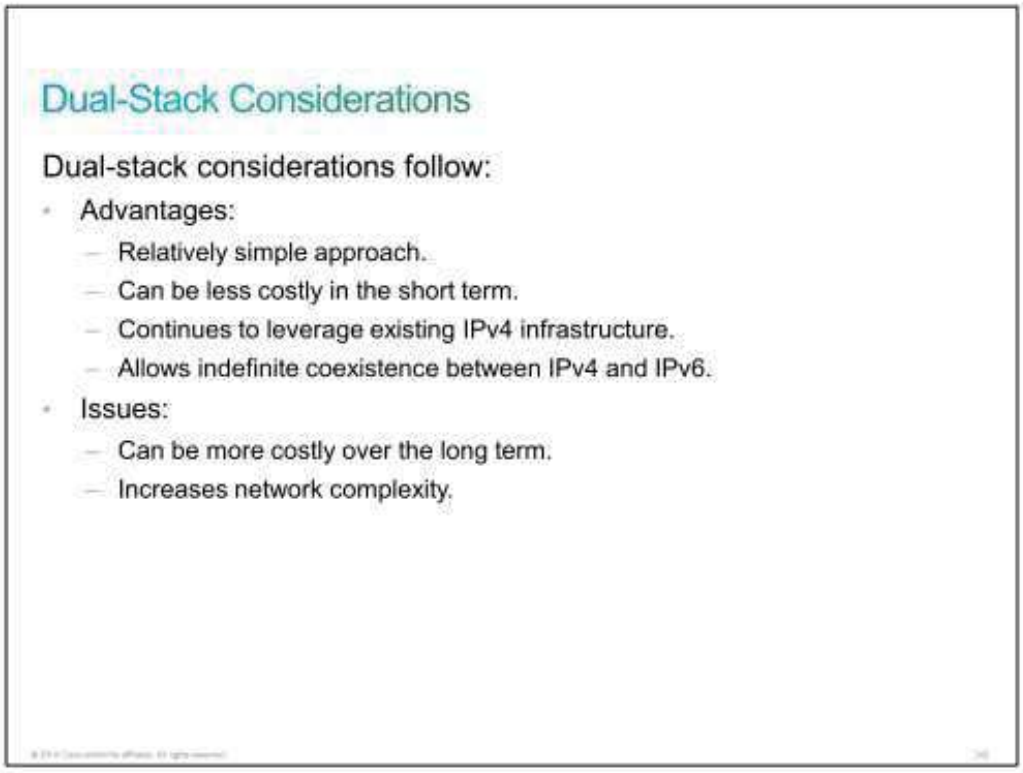
Because it is possible that the IPv6 stack can be misconfigured, some content providers (for example, Google) currently require that you register your domain to the content provider to receive DNS AAAA records in answers to queries. This requirement serves to prevent scenarios in which applications acquire IPv6 addresses, but an inoperative stack could result in no connectivity. In this case, it is better to use the IPv4 stack. In environments with working IPv6 implementations, the endpoints should receive AAAA answers and prefer the IPv6 stack for communication as intended.



https://t.me/learningnets

# Dual-Stack Considerations

In most cases, IPv6 integration requires dual stacking a node. The likelihood that all of the applications and services in a network will be IPv6-only is very small (at least for the foreseeable future). As a result, access to IPv4 connectivity will remain a requirement. However, every IPv6 integration scenario is different, requiring careful evaluation of the transition mechanisms that are available, including a review of the benefits and challenges of using any particular mechanism.




Dual stacking is a relatively easy approach to incorporating IPv6 into a network. Most modern operating systems support IPv6, so the majority of nodes (hosts, servers, or routers) probably support basic IPv6 enablement. This support results in a lower transition cost. Also, by dual stacking, an organization can continue to leverage its existing IPv4-only applications until those applications have been made IPv6-capable or have been phased out in favor of applications that support IPv6.

However, as with any environment where new technologies are introduced, the complexity of the environment will increase. It is not enough to simply IPv6 enable a desktop computer. Leveraging IPv6 requires additional infrastructure changes. These changes include modifications to services such as DNS, DHCP, security policies, and more. Supporting two protocols over time will presumably be more costly than only one protocol.

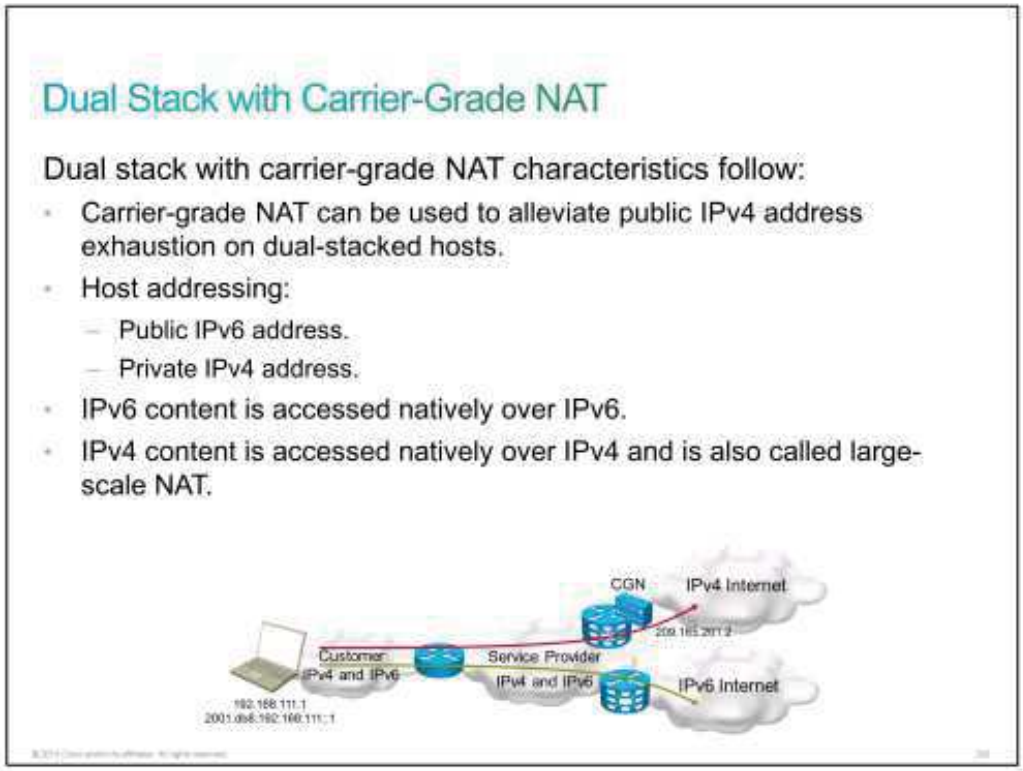
| | |
|---|---|
| **Note** | There is no standard way to transition a network to IPv6. The strategy depends on too many initial conditions, business plans, budget concerns, and so on. Each transition, or IPv6 integration into an existing network, is unique. |

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# Dual Stack with Carrier-Grade NAT

Some service providers soon will run out of assignable, public IPv4 addresses. The ultimate solution would be to migrate the Internet to IPv6 and assign IPv6 addresses to customers. However, until P-networks are on IPv6 and until all of the content is accessible on IPv6, there are several temporary solutions that can be used during the transition to IPv6.




One solution is CGN, also known as large-scale NAT. In CGN solutions, customers are assigned private RFC 1918 IPv4 addresses. These IP addresses are translated to public IPv4 addresses by NAT with a PAT device in the edge of the P-network, permitting the sharing of public IPv4 addresses among many customers. This shifts the NAT function and configuration from the customer premise to the P-network.

CGN can work with dual-stack hosts. Hosts are assigned public IPv6 addresses and private IPv4 addresses. When a host accesses IPv6 resources, the host accesses the resources over the IPv6 network natively. When the host accesses IPv4 resources, the host accesses resources over the IPv4 network. On the edge of the P-network, the host private IP address is translated into a public IP address, which is taken from the service provider pool (209.165.201.2, in the example).

| **Note** | The P-network should be enabled for dual stack to support this solution. |
|---|---|

Although the CGN principle is essentially the same as NAT in enterprise or even home environments, CGN introduces several challenges in terms of scalability and performance because a service provider often has to support hundreds of thousands or even more customers.

CGN has the following caveats:



https://t.me/learningnets

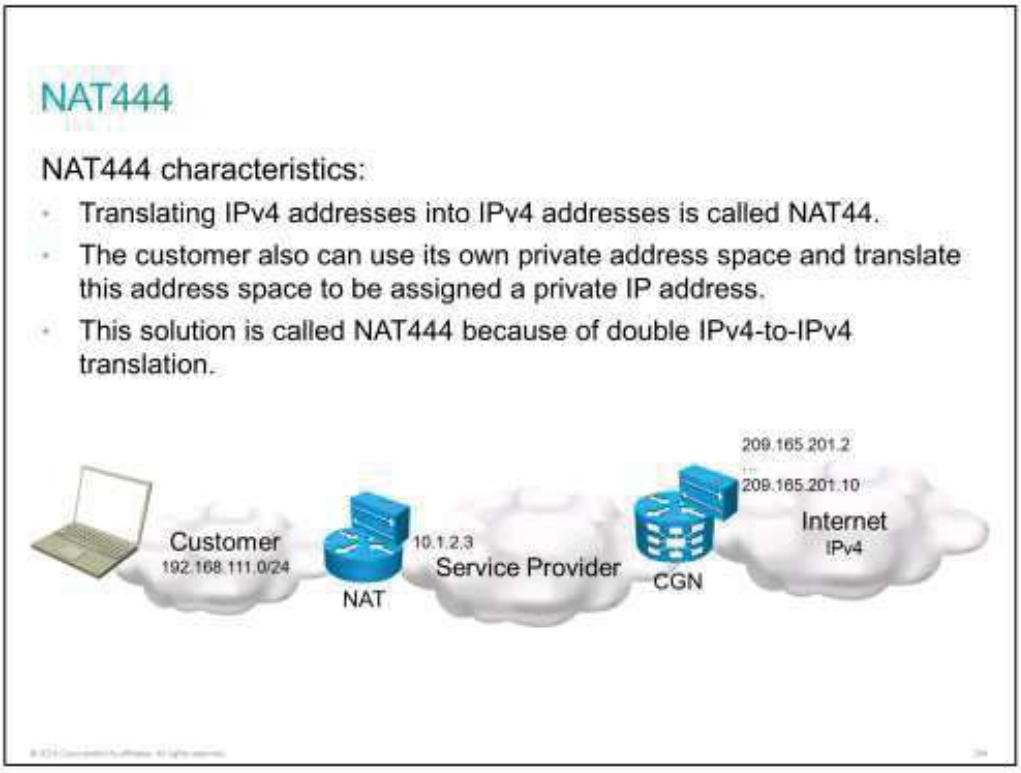
- It breaks end-to-end connectivity because one public IP address corresponds to many private IP addresses.
- It might have scalability and performance issues due to the stateful nature of NAT. Stateful nature means that a device performing NAT needs to track all translation across the device to successfully translate returning traffic.
- It makes record keeping operations more difficult because traffic from more customers seems to be originating from the same public IP address.
- It makes hosting of services on the customer side impossible because service providers will not configure static NAT with PAT for customers.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# NAT444

Translating one IPv4 address to another IPv4 address is called NAT44 because the packet traverses two IPv4 addressing domains.




If a customer uses its own private IPv4 address space and the service provider offers the CGN solution, the customer has to translate its private addresses to service provider-assigned private IPv4 addresses. This means that a packet traveling from the customer to the Internet is subject to network translation twice: once on the customer premises and once at the edge of the P-network. This solution is called NAT444 because a packet is translated twice and traverses three IPv4 addressing domains.

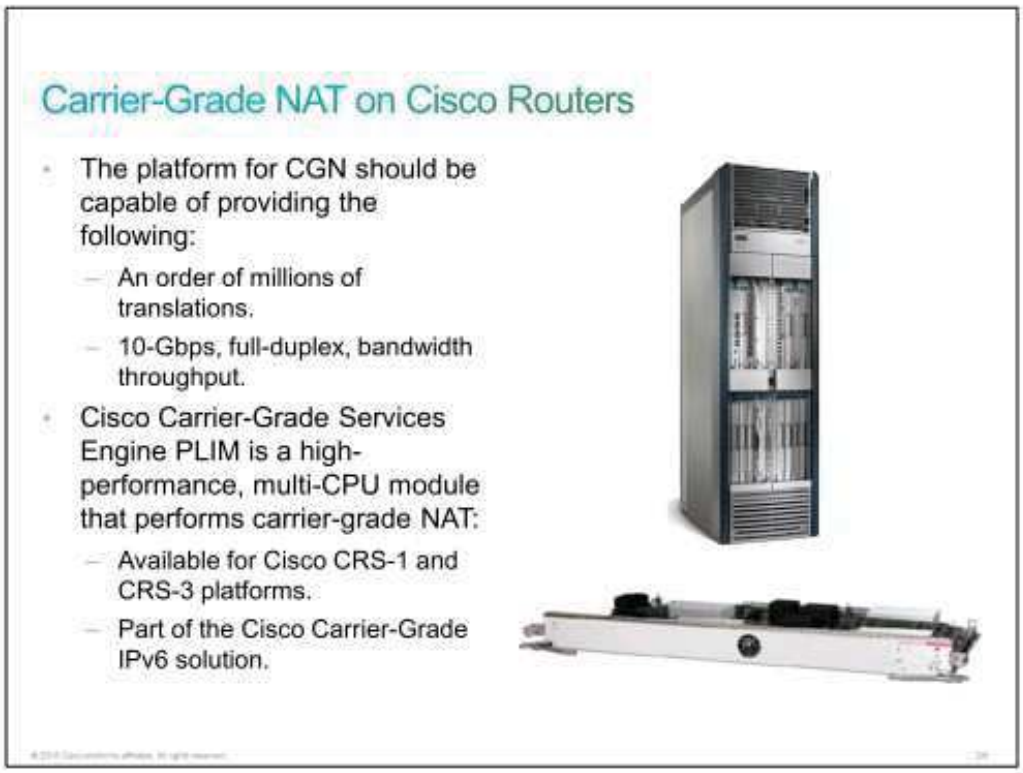
In the example, the customer uses the 192.168.111.0/24 private addressing space. However, the service provider assigned the 10.1.2.3 private IP address to the customer. The customer has to perform NAT with PAT first to translate from the 192.168.111.x address to 10.1.2.3. The service provider has to perform NAT with PAT again to translate from 10.1.2.3 to one of the public IP addresses (for example, 209.165.201.2).



https://t.me/learningnets

# Carrier-Grade NAT on Cisco Routers

As mentioned before, the CGN principle is essentially the same as NAT in enterprise or even home environments. However, a device performing CGN should be capable of providing an order of a million translations and 10 Gbps of full-bandwidth throughput in order to meet performance and scalability requirements of P-networks.




The Cisco solution that meets the CGN requirements is Cisco Carrier Routing System Router with Cisco Carrier-Grade Services Engine physical layer interface module.

The Cisco CGSE is an integrated multi-CPU service module offering carrier-class performance and scale in support of the Cisco Carrier-Grade IPv6 solution. The Cisco CGSE is a single-slot module that is supported on all models of the Cisco proven, high-end, carrier-class routing system: the Cisco CRS-1 and CRS-3 series. The Cisco CGv6 solution, running on one or more Cisco CGSE modules inside a Cisco CRS, can scale to tens of millions of IP address translations with tens of gigabits of performance to address IPv4 run-out and enable IPv6 transition. Several modules can be populated within a chassis for a high-performance solution that is deployable at places in the network where maximum Cisco CGv6 coverage can be obtained. The Cisco CGSE supports a highly available architecture with line-rate accounting and logging of translation information. Cisco IOS XR Software on the platform offers a flexible means to divert select packets through the Cisco CGSE, while enabling global IPv4 and IPv6 packets to traverse the Cisco CRS forwarding infrastructure as usual.

The Cisco CGSE interface module on the Cisco CRS offers service providers a near-term solution to address IPv4 run-out and preserve a service provider-present mode of operation. At the same time, it enables one or more methods to offer a low-risk, cost-effective means to activate IPv6 tunneling and translation functions.

The Cisco CGSE is supported on all of the form factors of the Cisco CRS-1 and CRS-3: 4, 8, and 16 slots and multichassis versions.



https://t.me/learningnets

The Cisco CGSE that is housed inside a Cisco CRS currently offers carrier-class performance for the following Cisco CGv6 services:

- Complete IPv4 and IPv6 routing and forwarding on the Cisco CRS platform.
- Service provider-class NAT44.
- Stateless NAT64, which will be discussed in the next topic.

The following are Cisco CGSE performance criteria:

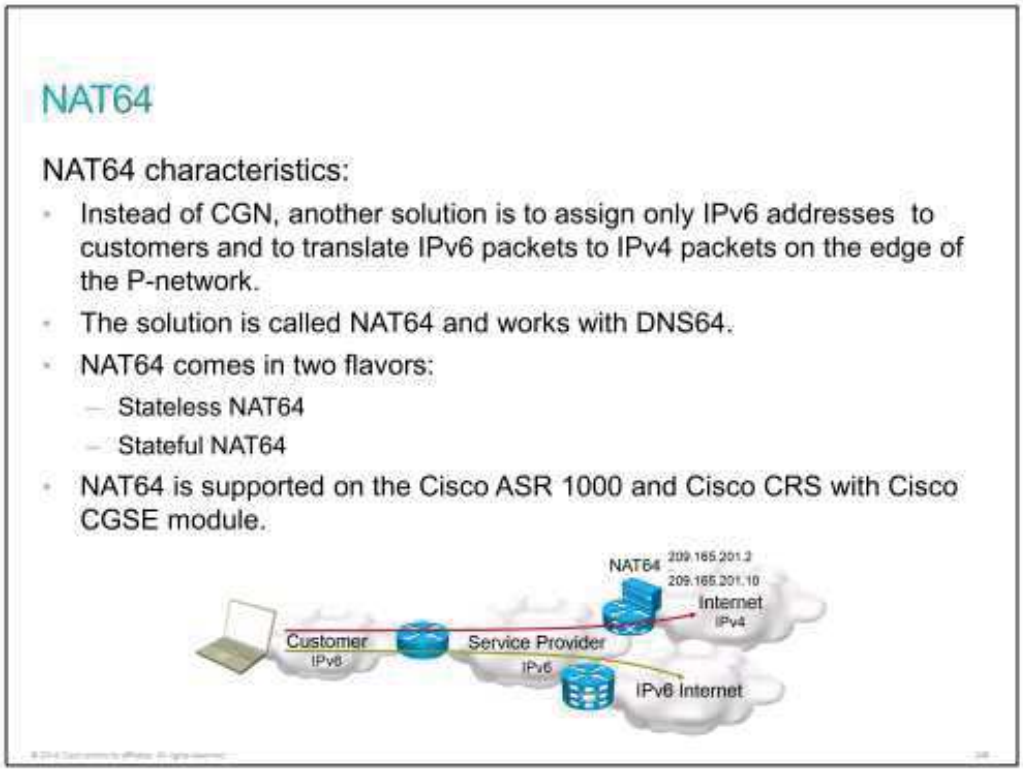
- 1+ million connection setups per second for stateful IPv4 NAT44.
- Line-rate forwarding for IPv4 and IPv6.
- Up to 20 million stateful NAT44 translations per Cisco CGSE module.
- Support for tens to hundreds of thousands of private IPv4 subscribers accessing the public IPv4 Internet.
- Ability to add multiple Cisco CGSE modules in a chassis, increasing performance linearly; maximum number of CGSE PLIMs per chassis is 4 slots—3, 8 slots—7, and 16 slots—12.



https://t.me/learningnets

# NAT64

As described in the previous topics, service providers soon will run out of assignable public IPv4 addresses. One of the solutions is CGN, where customers are assigned private IPv4 addresses. The second solution is to migrate the P-network to IPv6 and to assign IPv6 addresses to customers. Because all of the content is not available on IPv6, you will have to make sure that IPv6-only customers can access content on IPv4-only networks. Access of IPv4 content from IPv6-only hosts can be achieved by implementing NAT64 on the edge of the P-network, which takes IPv6 traffic on one side and converts it into IPv4 traffic on the other side and vice versa.




There are two types of NAT64:

- **Stateless NAT64:** Stateless NAT64 provides a stateless mechanism to translate an IPv4 header into an IPv6 header and vice versa using algorithmic binding between IPv4 and IPv6 addresses. Due to the stateless character, this mechanism is very effective. The disadvantage of stateless NAT64 is that it supports only one-to-one translation and is therefore not useful to battle against IPv4 address depletion.
- **Stateful NAT64:** Stateful NAT64 multiplexes many IPv6 devices into a single IPv4 address using PAT. The significant difference with stateful NAT64 is the elimination of the algorithmic binding between the IPv6 address and the IPv4 address. In exchange, state is created in the NAT64 device for every flow.

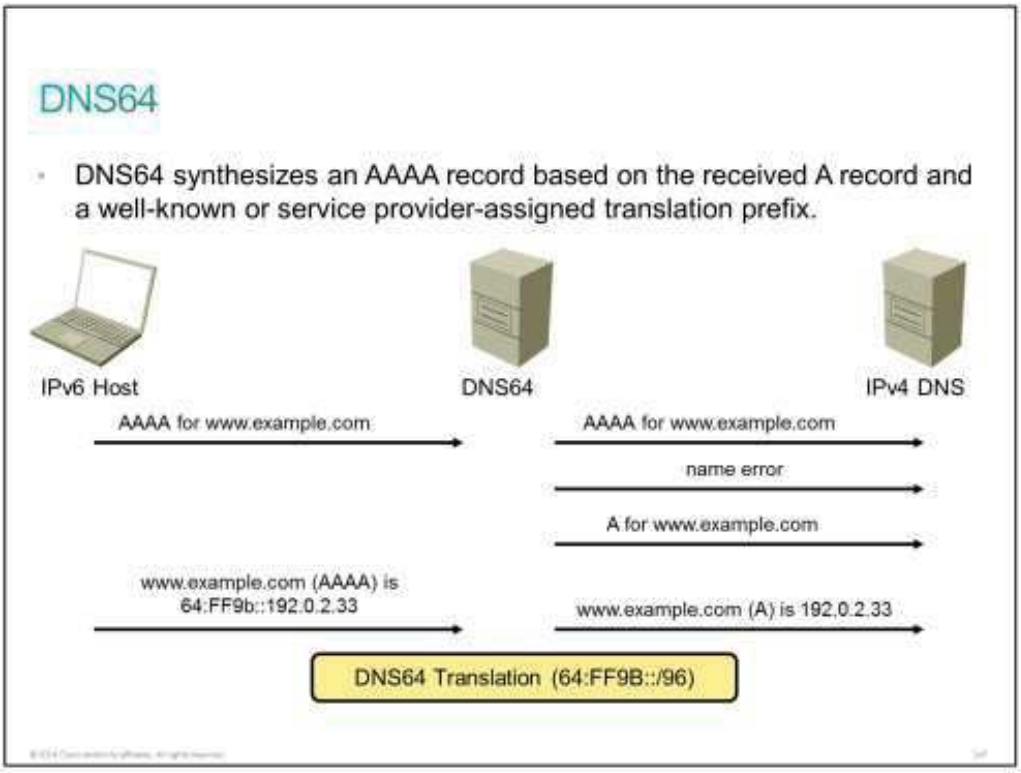
Both types of NAT64 work together with DNS64, which enables IPv6 hosts to retrieve IPv6 addresses of IPv4-only hosts.



https://t.me/learningnets

# DNS64

DNS64 is used with NAT64 and works as an addition to the standard DNS server. When DNS64 is asked for an AAAA record (IPv6 address for a hostname) but finds (or receives from an upstream DNS) only an A record, it synthesizes the AAAA record from the A record. The first part of the synthesized AAAA record is a well-known or service provider-assigned IPv6 prefix, which is also used with NAT64. The second part of the AAAA record is an IPv4 address of the IPv4 host that is found in the A record.




The WKP that has been assigned to IPv4 and IPv6 translation services is 64:FF9B::/96. However, an ISP also can use an NSP as part of its IPv6 address space as a translation prefix.

The IPv6 address for the synthesized AAAA record is combined by concatenating the translation prefix and IPv4 address. The IPv4 address immediately follows the translation prefix, and zeros are added to the IPv6 address if needed to fill all 128 bits in an IPv6 address. The following table presents concatenation of the IPv4 address with translation prefixes of different lengths.

| Prefix (NSP or WKP) | IPv4 Address | IPv4-Embedded IPv6 Address |
|---|---|---|
| 2001:db8::/32 | 192.0.2.33 | 2001:db8:c000:221:: |
| 64:ff9b::/96 | 192.0.2.33 | 64:ff9b::192.0.2.33 |

**Note** Notice that the IPv4 address can be written in hexadecimal form or in decimal dotted notation. The 192.0.2.33 address in the table can be written in hexadecimal as c000:0221.



https://t.me/learningnets

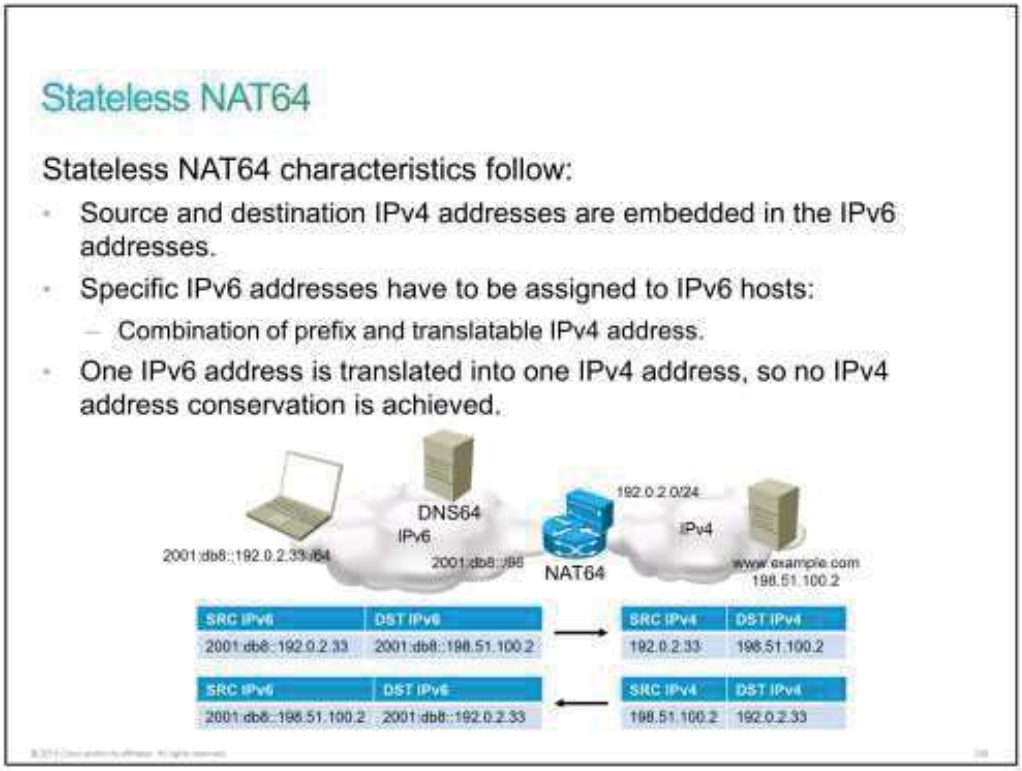
In the example, the IPv6 host wants to communicate with www.example.com. The host asks the DNS64 server for the AAAA record for www.example.com. DNS64 forwards the request to the authoritative DNS server. Because the www.example.com host is an IPv4-only host, the DNS server only has the A record for the host and answers to the DNS64 server with the name error. The DNS64 server then asks for the A record for www.example.com, and the DNS server returns it. The DNS64 server then concatenates the IPv4 address from the A record with a preconfigured translation prefix. In the example, the WKP is used and has a value of 64:ff9b/96. The DNS64 server then returns the IPv4-embedded IPv6 address to the IPv6 host inside the AAAA record. The concatenated IPv6 address in the example is 64:ff9b:: 192.0.2.33, which also can be written as 64:ff9b::c000:0221.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# Stateless NAT64

The key to the stateless NAT64 is in the fact that the IPv4 addresses are directly embedded in the IPv6 addresses. Limitations of stateless NAT64 translation are that it directly translates only the IPv4 options that have direct IPv6 counterparts, and it does not translate any IPv6 extension headers beyond the fragmentation extension header; however, these limitations are not significant in practice.




With a stateless NAT64, a specific IPv6 address prefix (as described with DNS64) will represent IPv4 systems within the IPv6 world. This range needs to be manually configured on the NAT64 device. Within the IPv4 world, all of the IPv6 systems have directly correlated IPv4 addresses that can be algorithmically mapped to a subset of the service provider IPv4 addresses. With this direct mapping algorithm, there is no need to keep state for any translation slot between IPv4 and IPv6. This mapping algorithm requires that the IPv6 hosts be assigned specific IPv6 addresses, using manual configuration or DHCPv6.

Stateless NAT64 will work successfully as proven in some of the largest networks, but it suffers from an important side effect: Stateless NAT64 translation will give an IPv6-only host access to the IPv4 world and vice versa, but it consumes an IPv4 address for each IPv6-only device that desires translation. Consequentially, stateless NAT64 is not a solution to the ongoing IPv4 address depletion. Stateless NAT64 is a good tool to provide Internet servers with an accessible IP address for both IPv4 and IPv6 on the global Internet. To aggregate many IPv6 users into a single IPv4 address, stateful NAT64 is required.



https://t.me/learningnets

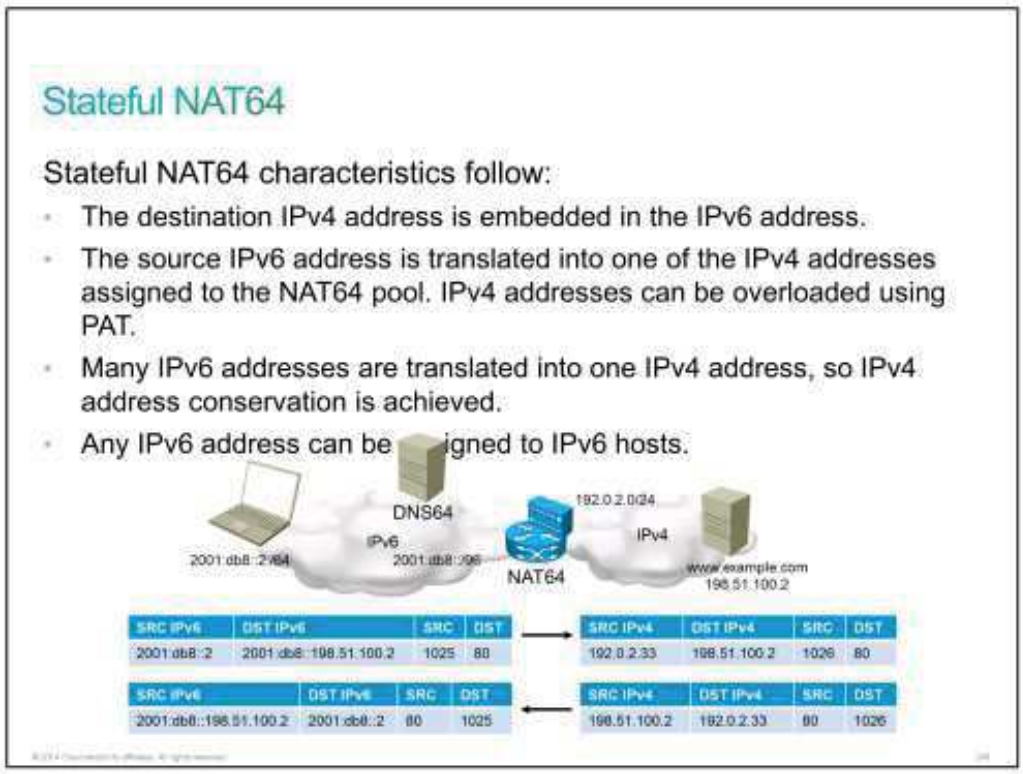
In the example, the IPv6-only host wants to communicate with the 198.51.100.2 IPv4-only host. The NAT64 device and DNS64 are configured with NSP 2001:db8::/96. The IPv4 subnet that is used to represent IPv6 addresses in the IPv4 network is 192.0.2.0/24. IPv6 hosts have to be preconfigured with IPv6 addresses that are a combination of the NSP and IPv4 subnet. In the example, the IPv6 host is assigned with the 2001:db8::192.0.2.33 IPv6 address. The destination IPv6 address that is received from the DNS64 server is 2001:db8::198.51.100.2. The IPv6 host creates an IPv6 packet with the previously mentioned IPv6 addresses. When the NAT64 device receives the packet, it translates the IPv6 header to the IPv4 header. IPv4 addresses are determined by stripping the translation prefix (2001:db8::/96) from the IPv6 addresses. NAT64 then sends the IPv4 packet toward the destination. When the NAT64 device receives the returning IPv4 packet, it translates the IPv4 header into the IPv6 header, and IPv6 addresses are determined by adding the translation prefix to the IPv4 addresses. The essence of stateless NAT64 is that addresses are algorithmically determined using the translation prefix, and there is no need to keep state for any translation slot across the NAT64 device.



https://t.me/learningnets

# Stateful NAT64

The significant difference with stateful NAT64 is the elimination of the algorithmic binding between the IPv6 address and IPv4 address. In exchange, state is created in the NAT64 device for every flow. Unlike stateless NAT64, stateful NAT64 does not consume a single IPv4 address for each IPv6 device that wants to communicate to the IPv4 Internet. More practically, this means that many IPv6-only users consume only single IPv4 addresses in a similar manner as IPv4-to-IPv4 NAT with PAT. This works well if the connectivity request is initiated from the IPv6 toward the IPv4 Internet. For example, if an IPv4-only device wants to speak to an IPv6-only server, manual configuration of the static translation slot will be required, making this mechanism less attractive to provide IPv6 services toward the IPv4 Internet.



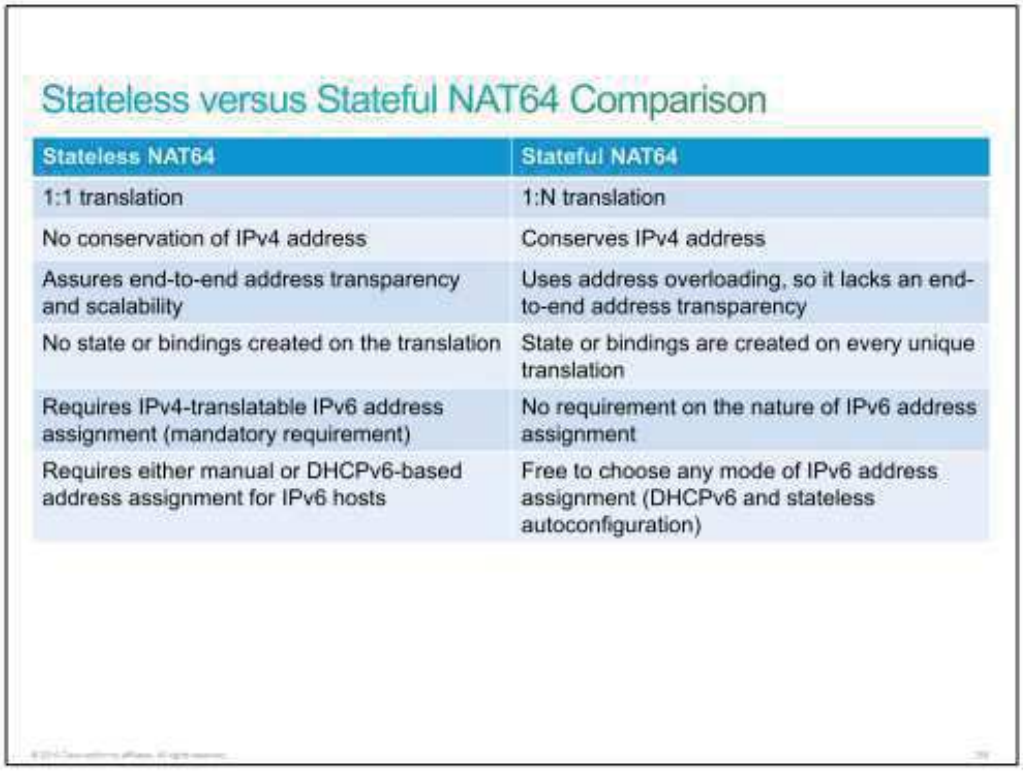

In the example, the IPv6-only host wants to communicate with the 198.51.100.2 IPv4-only host. The NAT64 device and DNS64 are configured with NSP 2001:db8::/96. The IPv4 subnet that is used to represent IPv6 addresses in the IPv4 network is 192.0.2.0/24 and is assigned to the NAT64 device as an IPv4 address pool. IPv6 hosts can have any IPv6 address that is configured by using manual configuration, DHCPv6, or stateless autoconfiguration. In the example, the IPv6 host is assigned with the 2001:db8::2 IPv6 address. The destination IPv6 address that is received from the DNS64 server is 2001:db8::198.51.100.2. The IPv6 host creates an IPv6 packet with the previously mentioned IPv6 addresses. When the NAT64 device receives the IPv6 packet, it translates the IPv6 header to the IPv4 header. The destination IPv4 address is determined by stripping the translation prefix (2001:db8::/96) from the destination IPv6 addresses. The source IPv4 address is chosen from the IPv4 pool and is used in the IPv4 packet. The source port of the original packet is also translated to enable overloading of a single IPv4 address. The NAT64 device creates an entry in the translation table that is used for subsequent and returning packets. The NAT64 device then sends the IPv4 packet toward the destination. When a NAT64 device receives the returning IPv4 packet, it translates the IPv4 header into the IPv6 header, and IPv6 addresses and ports are determined by performing a lookup in the translation table. The translated IPv6 packet is then sent toward the destination.



https://t.me/learningnets

# Stateless versus Stateful NAT64 Comparison

Stateless and stateful NAT64 are compared here:



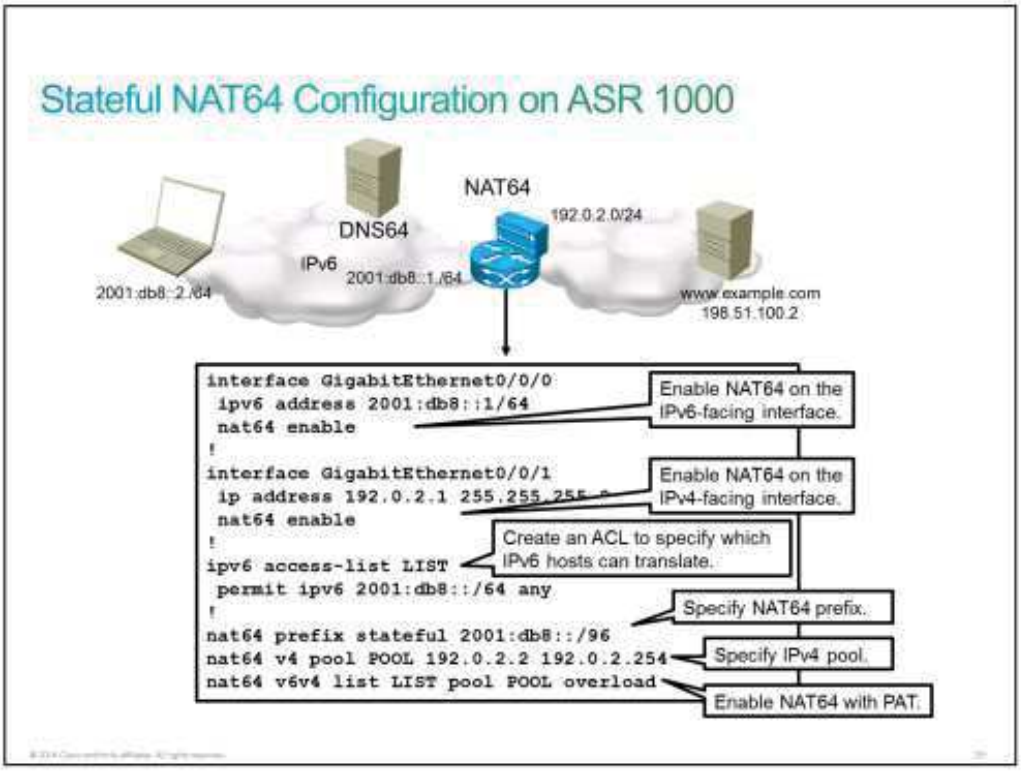

The table in the figure summarizes the differences between stateless and stateful NAT64.



https://t.me/learningnets

# Stateful NAT64 Configuration on ASR 1000

The stateful NAT64 configuration on the ASR 1000 is described here:




The figure shows a configuration example of stateful NAT64 configuration on the Cisco IOS XE Software platform. First, you have to enable NAT64 on IPv4 and IPv6-enabled interfaces using the **nat64 enable** command. Second, you have to specify hosts that are eligible for NAT64 translation, which is done by creating an IPv6 access list that permits hosts that can translate. Third, you have to specify the translation prefix using the **nat64 prefix stateful** command, followed by the translation prefix. To specify the IPv4 address pool, use the **nat64 v4 pool** command, followed by the pool name and starting and ending IPv4 address. In the example, IP addresses between 192.0.2.2 and 192.0.2.254 are assigned to the IPv4 address pool. Finally, bind together the access list that permits translation and the IPv4 address pool using the **nat64 v6v4 list pool** command with the optional **overload** keyword to enable PAT translation in addition to NAT.

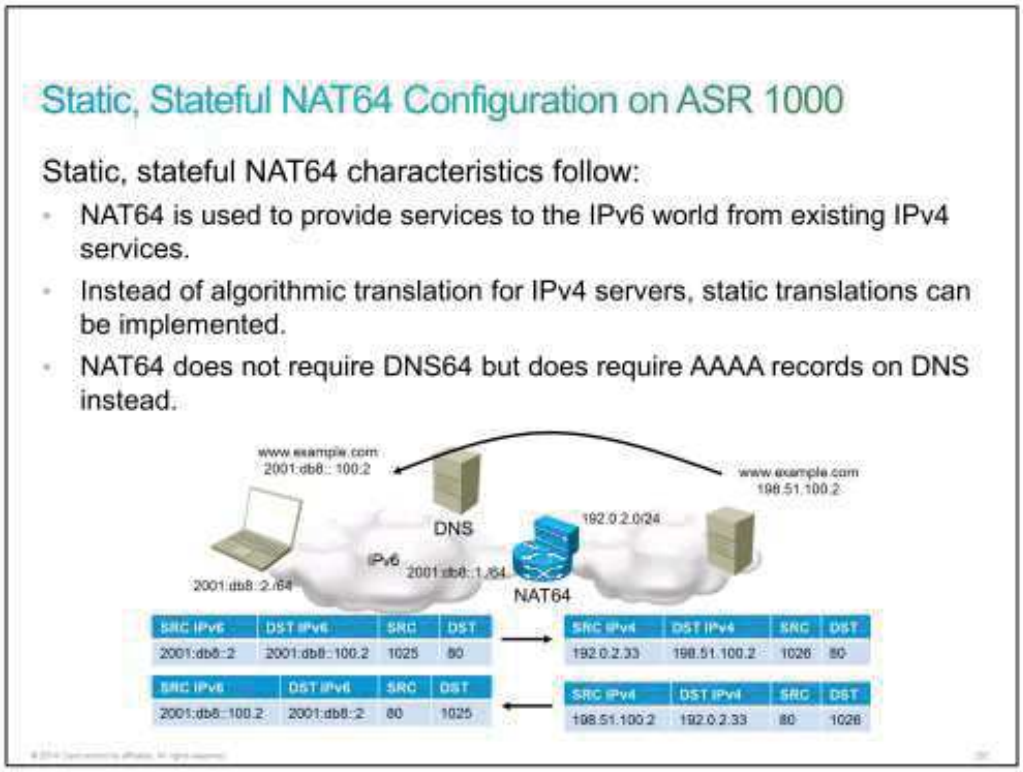
| | |
|---|---|
| **Note** | Configuration of NAT64 on Cisco routers (such as the Cisco ASR 1000 Series Aggregation Services Routers and Cisco Carrier Routing System) does not support a well-known translation prefix. Only NSP can be used. |
| **Note** | For a complete command reference, refer to the Cisco IOS XE Software Command Reference guides on www.cisco.com. |



https://t.me/learningnets

# Static, Stateful NAT64 Configuration on ASR 1000

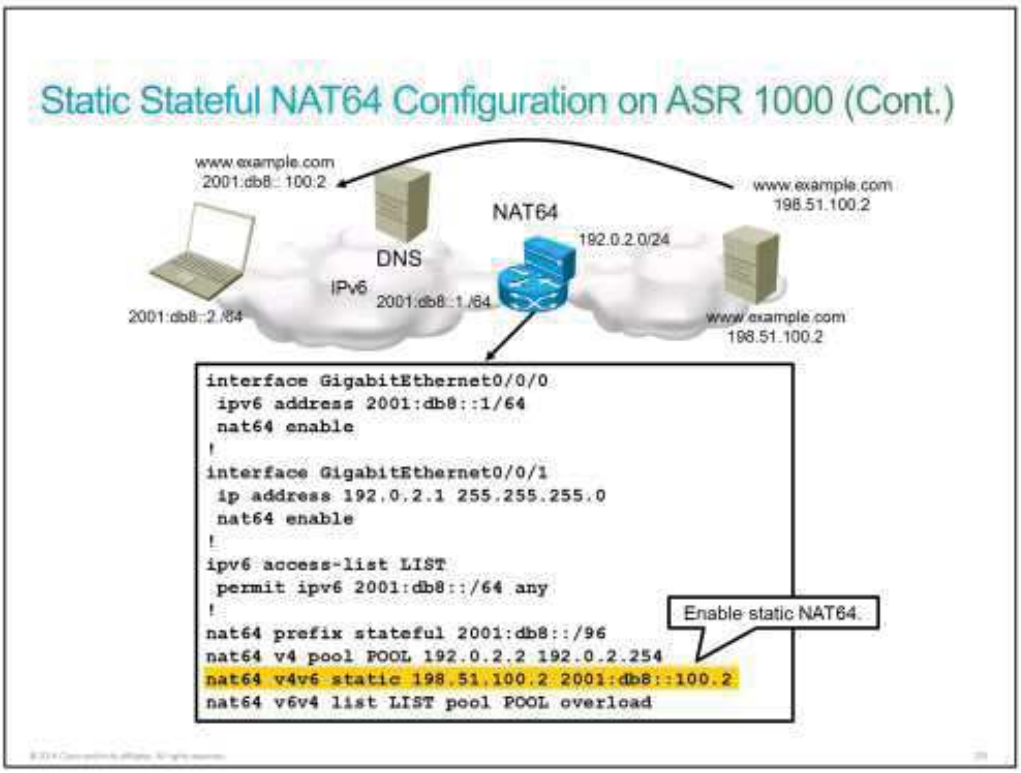
NAT64 is used to provide services from existing IPv4-only services to the IPv6-only world. When using stateful NAT64, the IP address of the IPv6-only host (that is, source IP address of the outgoing packet) is translated into one of the IPv4 addresses from the IPv4 pool. The IPv6 address of the IPv4-only server (that is, destination IP address of the outgoing packet) is still translated algorithmically using a translation prefix and DNS64. In a real-life scenario, it is desired that IPv4-only servers translate into predefined IPv6 addresses in the IPv6 domain. For this requirement, static NAT64 can be used. If there are static NAT64 translations, DNS64 is not required. The DNS server is responsible for providing IPv6 addresses of IPv4-only servers to the IPv6 domain.




In the example, the DNS server is configured with AAAA entry for www.example.com that points to 2001:db8:: 100:2. When the IPv6 host wants to communicate with the IPv4 server, it creates the IPv6 packet with the source IP address that is set to its own IPv6 addresses and the destination IPv6 address that is set to IPv6 addresses that are received from the DNS server. The NAT64 device translates the IPv6 header into the IPv4 header. The source IPv6 address is changed to the IPv4 address from the IPv4 address pool, and the destination IPv6 address is changed to the IPv4 address as configured with static NAT64. In the example, 2001:db8:: 100:2 translates into 198.51.100.2. For the returning packet, IPv4 addresses are translated back into IPv6 addresses. The source IPv4 address of the returning packet is translated back into the IPv6 address that is based on the static translation. The destination IPv4 packet is translated back into the IPv6 address that is based on the translation slot that was created for the outgoing packet.



https://t.me/learningnets




The figure shows a configuration example of a static, stateful NAT64 configuration on the Cisco IOS XE Software platform. In addition to the NAT64 configuration that was described before, another command is needed. Use the **nat64 v4v6 static** command, followed by the IPv4 address and IPv6 address to create static NAT64 translation. In the example, 198.51.100.2 is configured to translate into 2001:db8::100.2 and vice versa.

Similarly, you can use the **nat64 v6v4 static** command to statically translate IPv6 addresses of IPv6-only hosts to IPv4 addresses in the IPv4 domain.

An example follows:

```
nat64 v6v4 static 2001:db8::2 192.0.2.2
```

The IPv6 address 2001:db8::2 is statically mapped to IPv4 address 192.0.2.2



https://t.me/learningnets

## Static, Stateful NAT64 Configuration on ASR 1000 (Cont.)

- Displays NAT64 translations:

```
NAT64# show nat64 translations
Proto  Original IPv4       Translated IPv4
       Translated IPv6     Original IPv6
------------------------------------------------------------------------------
---    198.51.100.2        2001:db8::100.2
       ---
icmp   198.51.100.2 :1     [2001:db8::100.2]:6520
       192.0.2.2 :1        [2001:db8::2]:6520
Total number of translations: 2
```

Use the **show nat64 translations** command to verify NAT64 translations. The command displays static and dynamic translations.

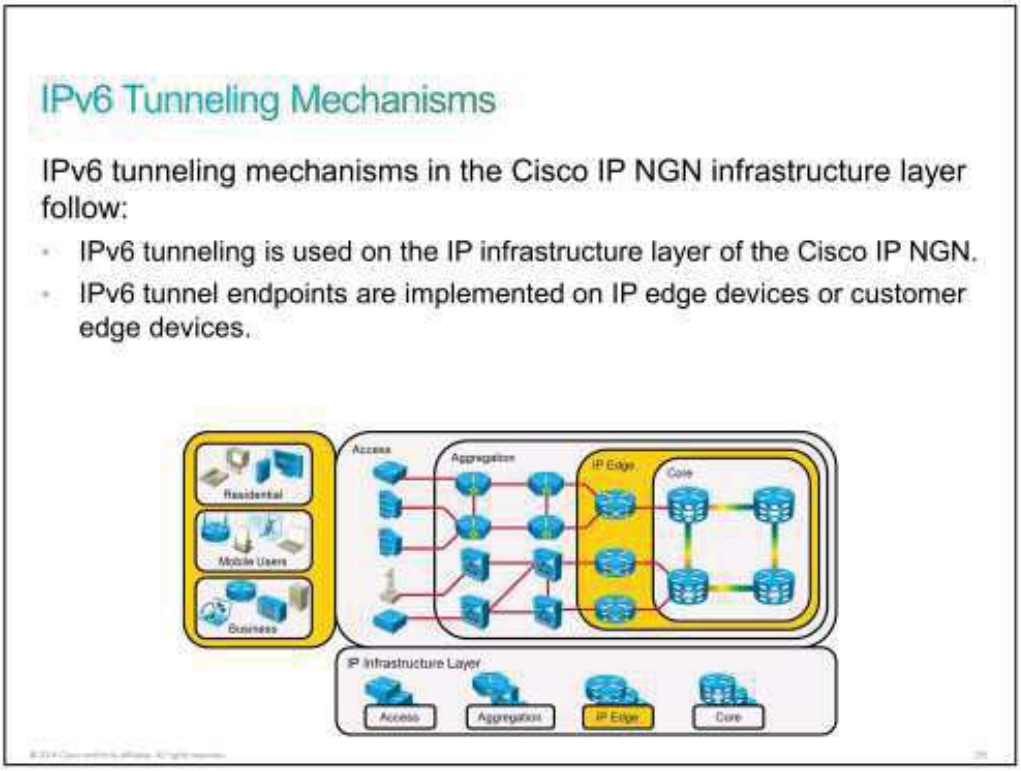
In the example, the output is a result of the configuration from the previous figure when the IPv6 host at 2001:db8::2 pings the IPv4 server at 2001:db8::100.2. Translation between IPv4 and IPv6 addresses of the IPv4 server is configured statically as shown in the previous figure. The first entry shows the static translation and is permanent, while the second entry is built dynamically when a packet traverses the NAT64 device. The second entry shows that the source IPv6 address of the IPv6 host is translated into one of the IPv4 addresses from the IPv4 address pool, while the destination IPv6 address is translated statically, as also shown in the first entry.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# IPv6 Tunneling Mechanisms

IPv6 tunneling is used on the IP infrastructure layer of the Cisco IP NGN.




IPv6 tunneling can be implemented on service provider IP edge devices if customer devices are not managed. If customer devices are managed, tunneling also can be implemented on CE devices. This way, the service provider can have the entire network be IPv4-only. Only CE devices and relay routers should be configured as dual-stack devices.

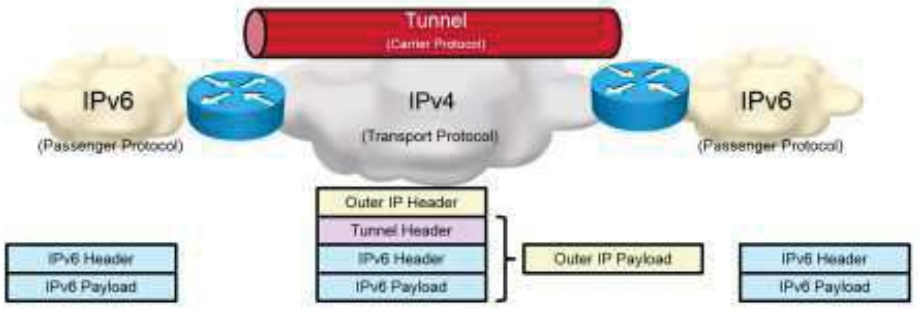


https://t.me/learningnets

## IPv6 Tunneling Mechanisms (Cont.)

IPv6 tunneling mechanism characteristics follow:

- Tunneling is used to transport one network protocol over another by encapsulating packets.
- During the transition, not all devices could be configured with dual stack:
  - Not all devices are under common administration.
  - There is almost twice as much of an administrative burden as with single-stack networks.




Tunneling is used to connect two disjointed networks that do not have a native routing path to each other. Protocols that are used in these disjointed networks are called passenger protocols. Tunneling is done using a routing protocol across a transport network. This example of tunneling is the connection of two disjointed IPv6 networks over an IPv4 network. The IPv6 networks do not have direct connectivity, but they can communicate over the tunnel across the IPv4 network.

In IP tunneling, the entire IP packet, including the header and payload, is encapsulated within another packet that is native to the transport network. Depending on the tunneling protocol, an additional tunnel header is introduced between the original (passenger) and transport header. The transport header is also called the outer header. The routing of an encapsulated packet across the transport header is performed based on the outer header.

At the borders between the source passenger network and the transport network as well as the transport network and the destination passenger network, routers are used as the gateways. To the packets that are traversing the gateway from the source network, the tunneling header and the outer IP header are added. The tunneling header and the original packet become the outer IP payload. When the packet traverses the gateway between the transport and destination network, the gateway strips the outer and tunneling headers from the packet and sends the original packet to the destination network.

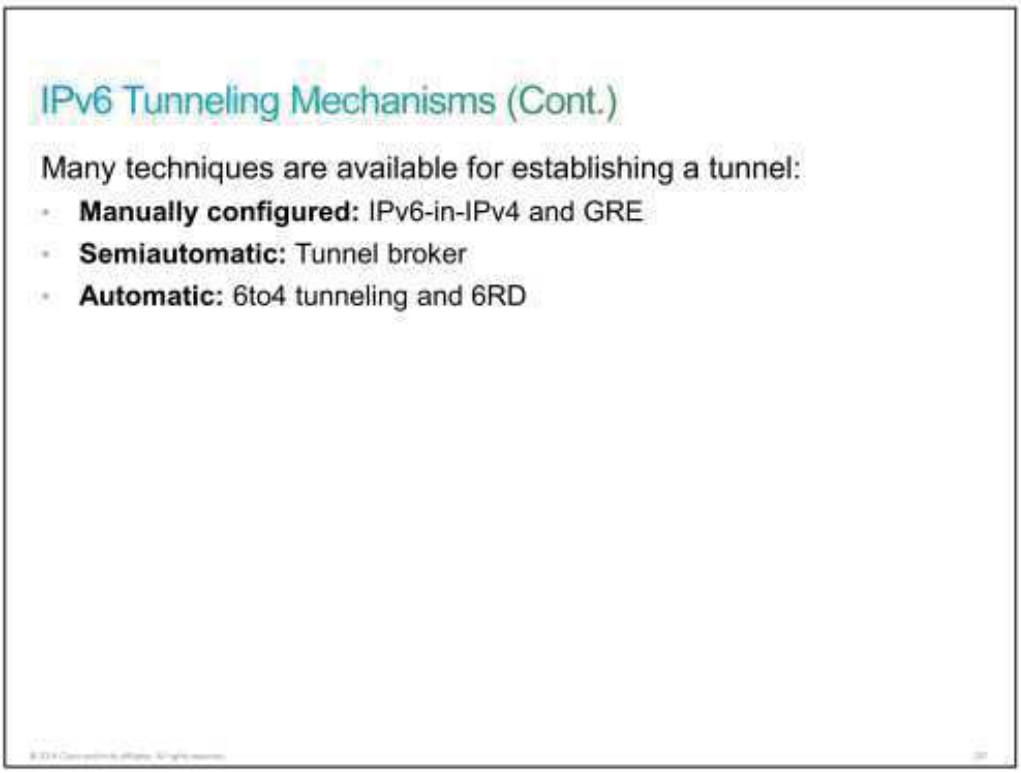
The most straightforward approach to transition to IPv6 is to enable dual stack on all devices on the Internet. However, this approach has several limitations:

- Not all devices are under the administration domain, so dual stack cannot be enabled on all devices at the same time.
- Dual stack requires almost twice as much administrative burden as single-stack networks because addressing and routing have to be configured and monitored for each stack separately.

Because of these limitations, the result is islands of IPv6 and IPv4 networks. These networks can be interconnected using tunneling technologies.

 © 2014 Cisco Systems, Inc.

https://t.me/learningnets




Many techniques are available to establish tunnels between IPv6 islands:

- **Manual tunneling between two or more sites:** With manual tunneling, one tunnel has to be configured to interconnect the two IPv6 sites. To connect to the IPv6 Internet, the tunnel has to be configured to tunnel brokers. Because tunnels have to be provisioned manually, this solution does not scale well. To encapsulate IPv6 packets inside IPv4, 6in4 encapsulation or GRE can be used.
- **Semiautomatic tunneling between a site and the rest of the Internet:** IPv6 tunnel brokers provide IPv6 tunnels to connect a site with the IPv6 Internet.
- **Automatic tunneling between sites and the rest of the Internet:** With dynamic tunneling, tunnels between IPv6 islands are established automatically. When two IPv6 sites want to communicate, the IPv6 destination address will contain the embedded IPv4 tunnel destination address, and a tunnel will be established automatically. To connect to the IPv6 Internet, tunnels have to be established to the "relay routers" that offer connectivity to the IPv6 Internet. There are two technologies that are available for automatic tunneling between IPv6 sites: 6to4 and 6RD.

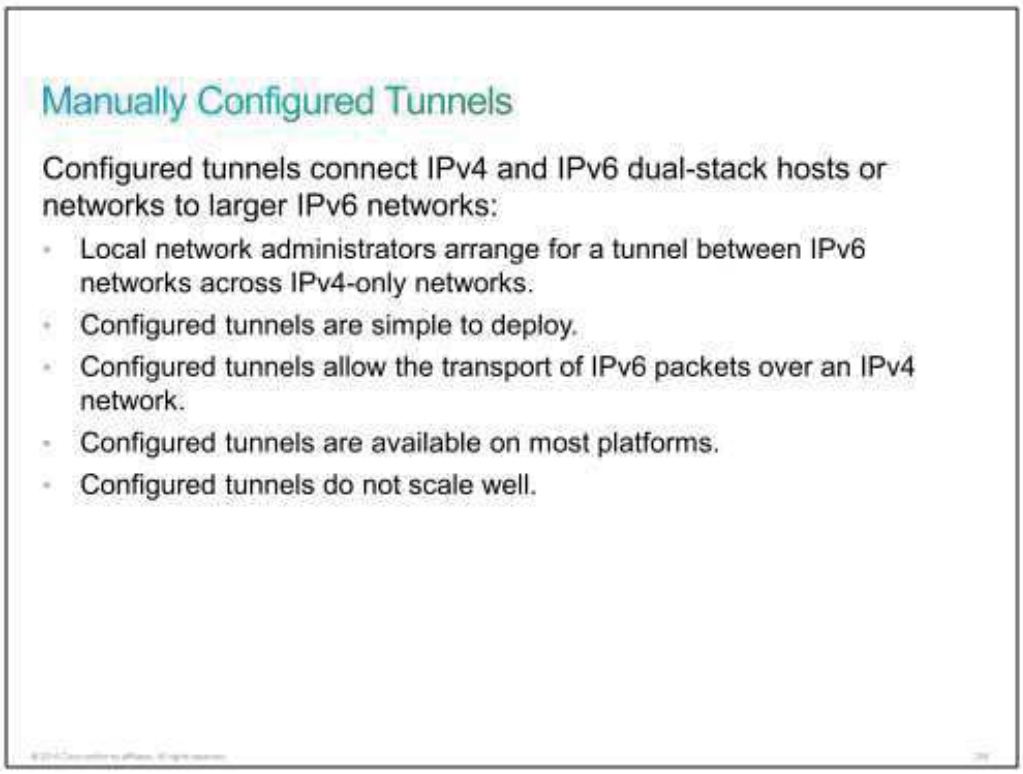


https://t.me/learningnets

# Manually Configured Tunnels

Manually configured tunnels have a few advantages over automatic tunnels. Most significantly, from a security point of view, static tunnels are not activated without administrator knowledge. A local administrator must explicitly configure a tunnel between two points.

These tunnels are also simple to deploy because one needs to set up just two endpoints (or gateways), configure the IP addresses to which the tunnel is attached, and enable the routing between the two endpoints.




Simple static tunnel support is also available on most platforms because extra processing is not required.

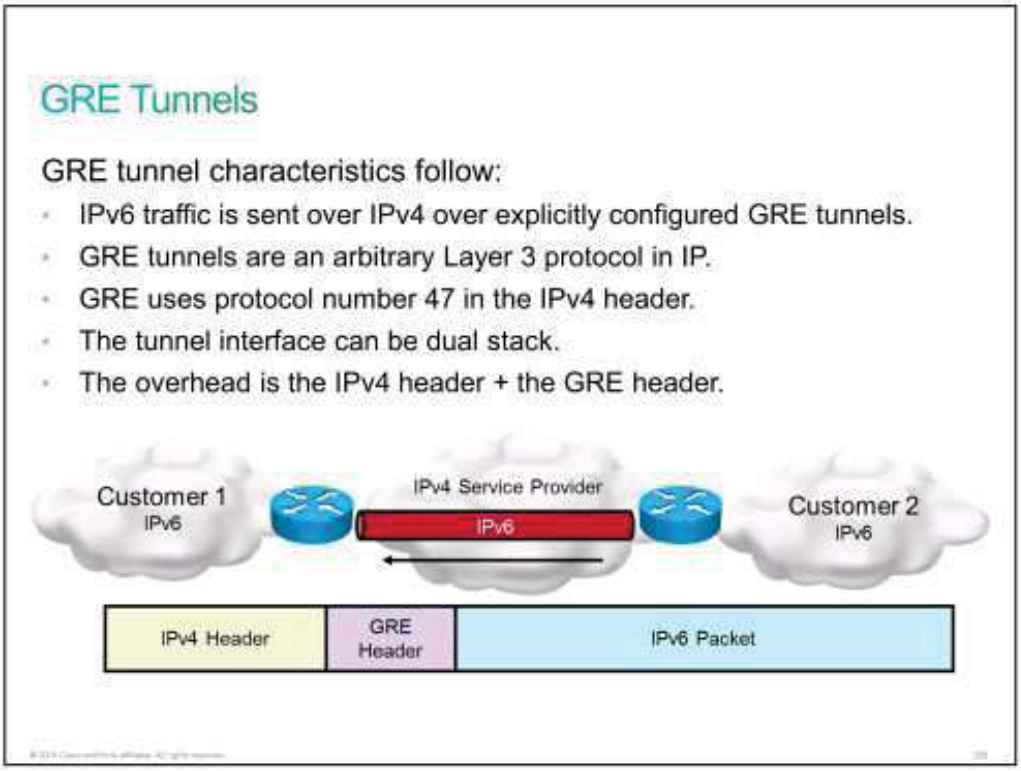
The disadvantage of static tunnels is that they do not scale well. If a full mesh of tunnels is required, the number of tunnels will grow exponentially to (n-1)/n.



https://t.me/learningnets

# GRE Tunnels

Another encapsulation method that can be used to tunnel IPv6 packets inside IPv4 is GRE.



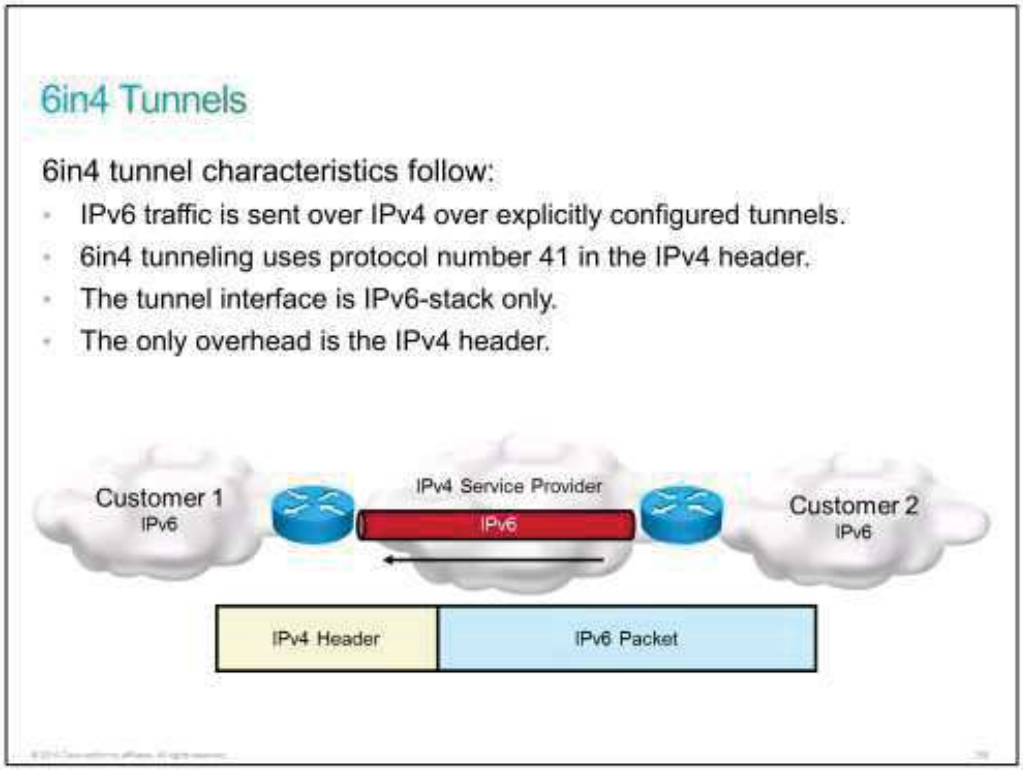

GRE is a tunneling protocol that can tunnel an arbitrary Layer 3 protocol inside IPv4 or IPv6. For IPv6 transition, tunneling of the IPv6 protocol inside IPv4 can be accomplished using GRE. GRE uses protocol number 47 inside the IPv4 packet. As opposed to 6in4 tunneling, GRE introduces additional GRE overhead, which is 20 B long. As mentioned before, the 6in4 tunnel is IPv6-stack only, while the GRE tunnel can be dual stack and therefore used to tunnel both IPv4 and IPv6 packets.



https://t.me/learningnets

# 6in4 Tunnels

6in4 is an encapsulation method that can be used in manually provisioned tunnels.



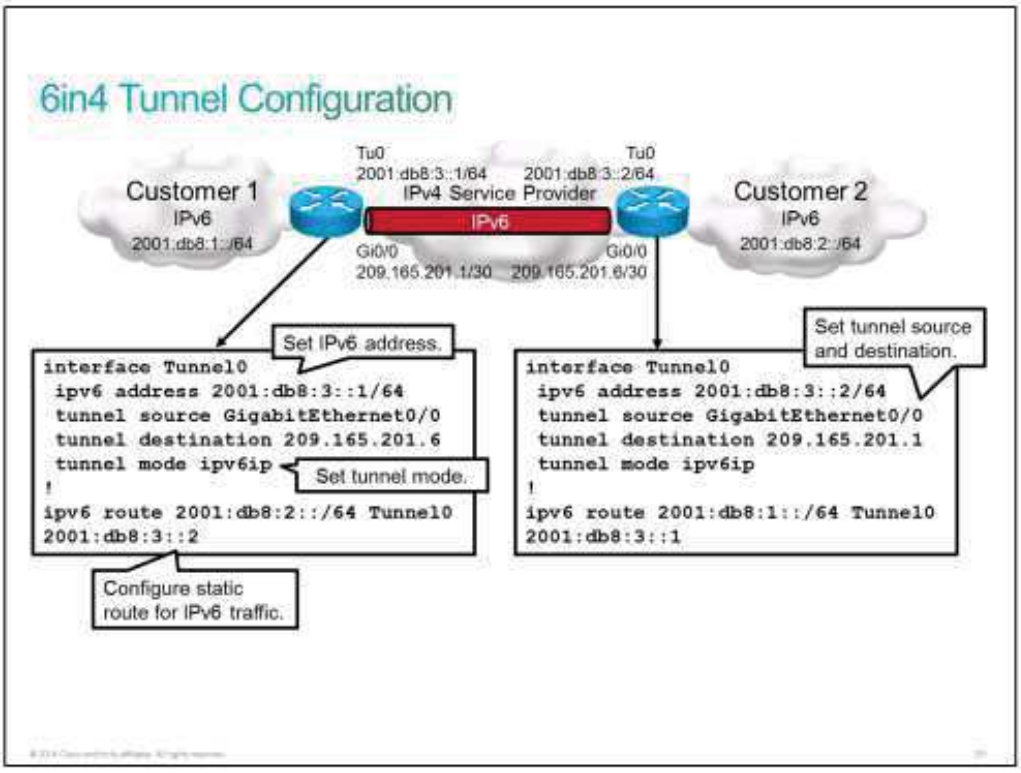

To connect two sites, one tunnel has to be configured. To connect to another site, two additional tunnels have to be configured from both sites to the third one. 6in4 tunneling uses protocol number 41 in the IPv4 header. The only overhead that is introduced to encapsulated IPv6 packets inside an IPv4 packet is the IPv4 header (at 20 B). The tunnel interface that uses 6in4 encapsulation is IPv6-stack only and can be used to tunnel IPv6 packets only.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# 6in4 Tunnel Configuration

The 6in4 tunnels are supported only on Cisco ISR Series routers and on Cisco ASR 1000 Series routers. Therefore, configuration for Cisco IOS and IOS XE Software is shown.




To set up a tunnel, first create a tunnel interface using the **interface tunnel** command, followed by the interface identifier. A tunnel interface is a virtual interface that acts the same as a physical interface. The tunnel interface requires an IP address, and traffic can be routed across a tunnel interface. To tunnel IPv6 traffic across the tunnel interface, you have to set an IPv6 address using the **ipv6 address** command. It is important that both ends of the tunnel have an IP address from the same network because tunnel interfaces should seem to be directly connected. Then you have to use the **tunnel source** command to specify the tunnel source, which is a physical interface that acts as a tunnel endpoint on the local router. You have to use the **tunnel destination** command to specify a tunnel destination as well, which is an IP address of a tunnel endpoint on the router that is terminating the tunnel. Because many types of tunnels are supported on Cisco routers, you have to specify the tunnel type by using the **tunnel mode** command. To create a 6in4 tunnel, the tunnel mode has to be set using the command **ipv6ip**. Finally, you have to route traffic for the destination IPv6 network over the tunnel interface by using the **ipv6 route** command, where the outgoing interface is set to the tunnel interface and the next hop is set to the IPv6 address of the tunnel interface on the other side of the tunnel.

Configuration of the GRE tunnel is similar to the 6in4 tunnel. The only difference is that the **tunnel mode** command should be set using the **gre ip** command. The following is a sample configuration for GRE encapsulation:

```
interface Tunnel0
 ipv6 address 2001:db8:3::1/64
 tunnel source GigabitEthernet0/0
 tunnel destination 209.165.201.6
 tunnel mode gre ip
```



https://t.me/learningnets

| Note | For a complete command reference, refer to the Cisco IOS and IOS XE Software Command Reference guides on www.cisco.com. |
|---|---|

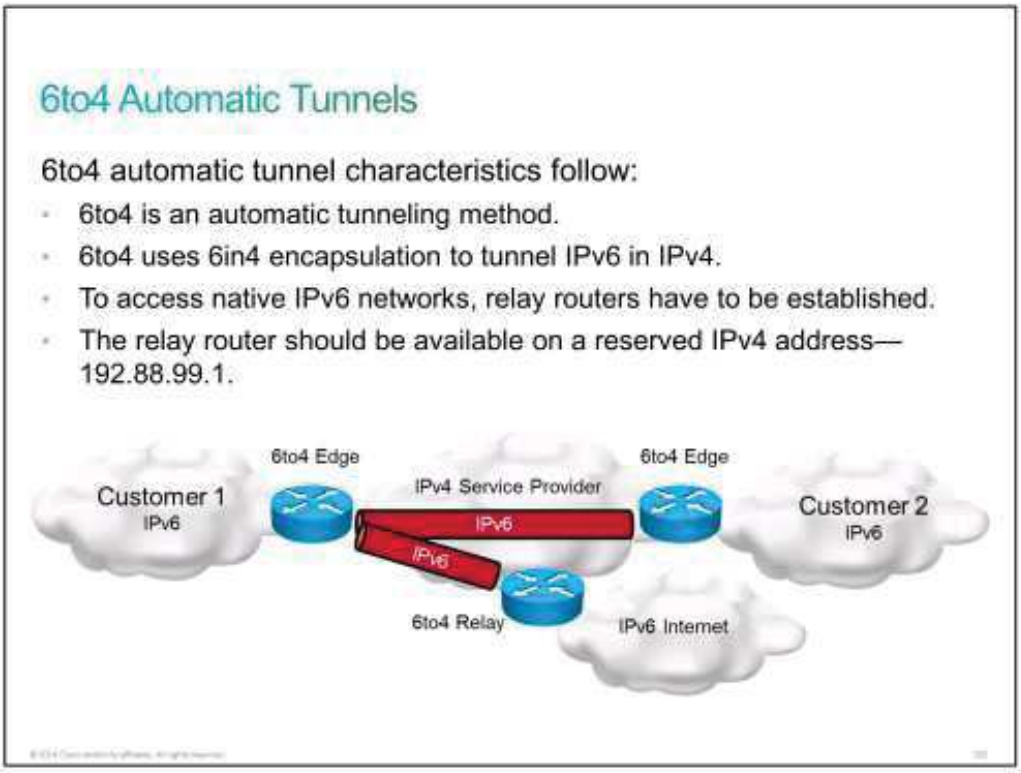
GRE tunnels are also supported on Cisco IOS XR Software platforms. However, tunneling of IPv6 over IPv4 using GRE tunnels is not supported on Cisco IOS XR Software platforms.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# 6to4 Automatic Tunnels

The 6to4 tunnel method is an automatic tunnel establishment method that enables the connection of IPv6 islands through an IPv4 network. 6to4 uses 6in4 encapsulation to tunnel IPv6 in IPv4. It also gives a valid IPv6 prefix to each IPv6 island. This prefix enables the fast deployment of IPv6 in a corporate network without the need to access addresses from ISPs or registries because the 2002::/16 IPv6 address block is specifically reserved for 6to4 tunneling.




The term 6to4 "edge router" refers to the 6to4 router that connects to the customer IPv6 network. The term 6to4 "relay router" refers to the 6to4 router that connects to the IPv6 Internet.

The 6to4 tunnel method enables the edge router to forward packets to any destination with a 2002::/16 prefix. However, all other destinations are unreachable unless one of the 6to4 edge routers, called a 6to4 relay, offers traffic forwarding to the IPv6 Internet. In this case, all other 6to4 edge routers include a default route entry to this 6to4 relay. The 6to4 relay router should be available on the reserved 6to4 relay IPv4 address. The 192.88.99.0/24 address block is allocated for use as the 6to4 relay anycast addresses according to RFC 3068. The anycast address of 192.88.99.1 has been allocated for sending packets to a 6to4 relay router.



https://t.me/learningnets

## 6to4 Automatic Tunnels (Cont.)

6to4 addressing characteristics follow:

- IPv6 networks have to use specially assigned prefixes—2002::/16.
- The well-known prefix is concatenated with the IPv4 address that is assigned to the customer. The resulting IPv6 network is assigned to the customer.
- Traffic between IPv6 islands: The tunnel destination is determined from the IPv6 destination prefix.
- Traffic to the IPv6 Internet: The 6to4 relay router IPv4 address is known.




6to4 requires a special code on the edge routers, but the IPv6 hosts and routers inside the 6to4 site do not require any new features to support 6to4. Each 6to4 site receives a /48 prefix that concatenates the 2002 and the IPv4 address of the edge router. For example, if the IPv4 address of the edge router is 209.165.201.1, the prefix of its IPv6 network will be 2002:*IPv4 address of the edge router*::/48, which is 2002:d1a5:c901::/48 (because d1a5:c901 is the hexadecimal representation of 209.165.201.1). The IPv6 network then can subnet this space, as with any /48 prefix.

When an IPv6 packet goes to another IPv6 island, the destination address is in the 2002::/16 range. When this packet reaches the 6to4 edge router, the 6to4 edge router extracts the IPv4 address that is embedded in the 2002:: destination address. The 6to4 edge router then will encapsulate the IPv6 packet in an IPv4 packet, with the IPv4 destination address extracted from the IPv6 destination address. This IPv4 address represents the address of the other 6to4 edge router of the destination 6to4 site. The destination edge router de-encapsulates the IPv6 packet in the IPv4 packet and then forwards the native packet toward its final destination.

When an IPv6 packet goes to the IPv6 Internet, the default route that points to the 6to4 relay will be used.

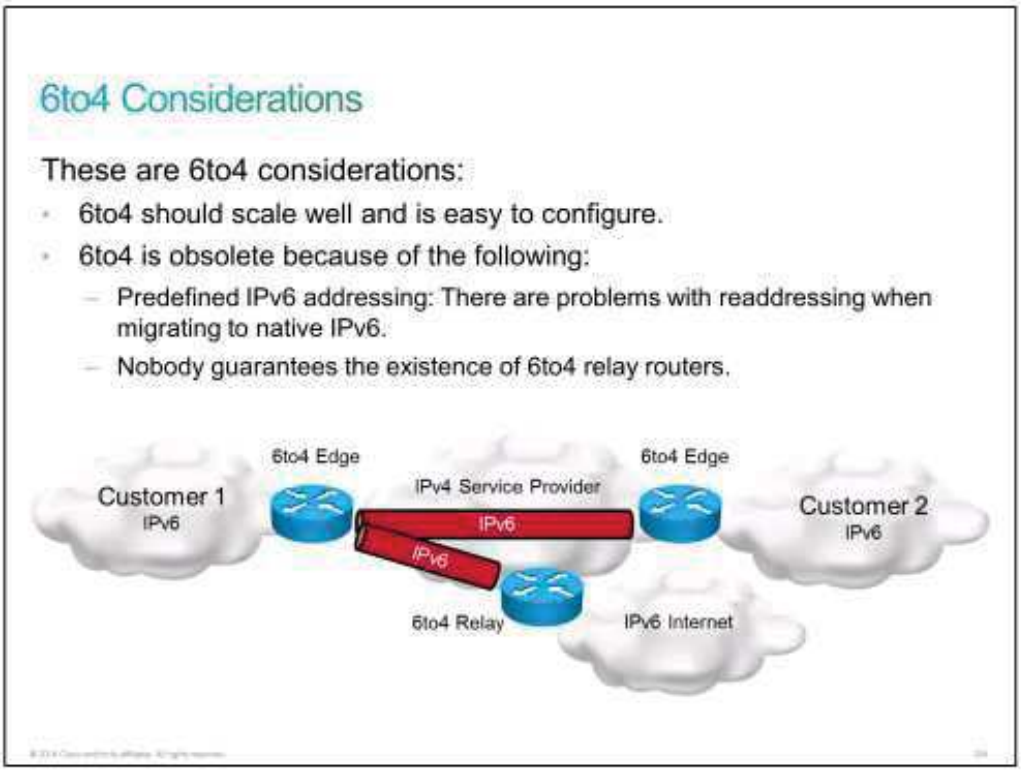
If IPv4 protocol 41 ("IPv6-in-IPv4 encapsulation") is filtered anywhere along the path, the tunnel will not work.



https://t.me/learningnets

# 6to4 Considerations

6to4 tunneling should scale well and should be easy to configure due to automatic establishment of IPv6 tunnels.




There are two significant issues with 6to4 that made it obsolete:

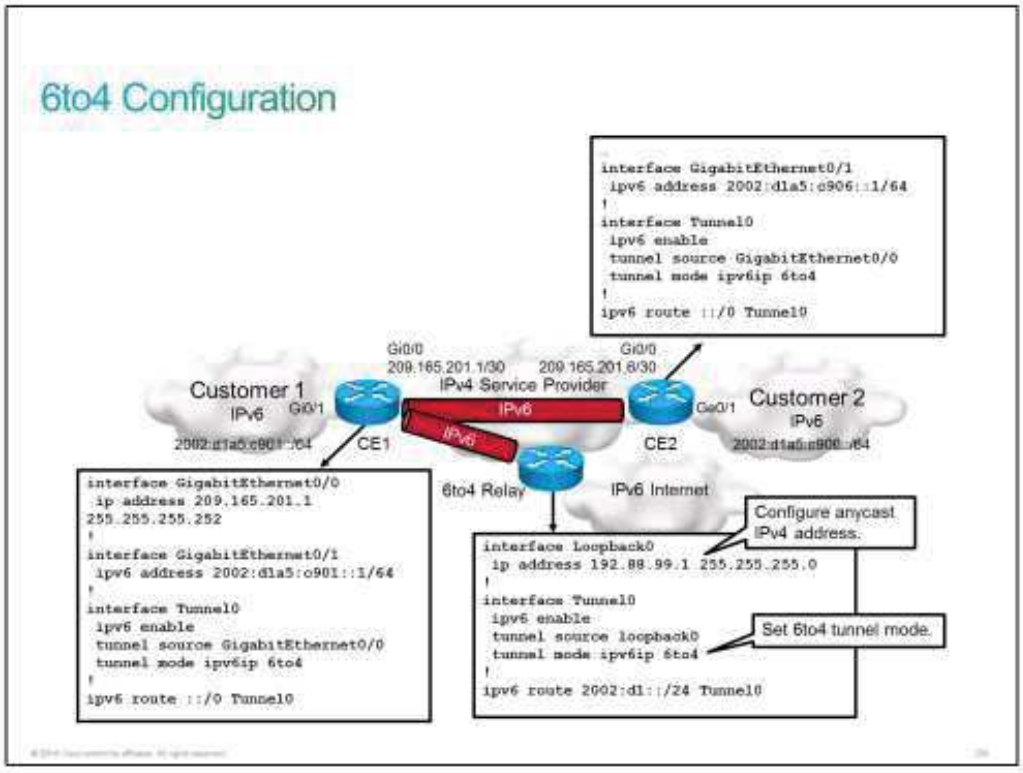
- **Predefined IPv6 addressing:** If the IPv4 address of the 6to4 edge router changes, the entire site /48 self-allocation also changes, and the 6to4 environment must be renumbered. If the customer migrates to native IPv6 access, new IPv6 addressing will be assigned to the customer, and the customer network will have to be renumbered again.
- **Existence and maintenance of 6to4 relay routers:** No one is responsible to set up and maintain the 6to4 relay routers, and nobody guarantees that 6to4 relay routers will be reachable.



https://t.me/learningnets

# 6to4 Configuration

6to4 tunnels are supported only on Cisco ISR Series routers and on Cisco ASR 1000 Series routers. The configuration for Cisco IOS and IOS XE Software is shown here:




To configure 6to4 on the CE router, first create a tunnel interface using the **interface tunnel** command, followed by the interface identifier. Enable the tunnel interface for IPv6 using the **ipv6 enable** command, which will configure link-local IPv6 addresses on the tunnel interface. Then use the **tunnel source** command to specify the tunnel source, which is a physical interface that acts as a tunnel endpoint on the local router. Because many types of tunnels are supported on Cisco routers, you have to specify the tunnel type by using the **tunnel mode** command. To create a 6to4 tunnel, the tunnel mode has to be set using the **ipv6ip 6to4** command. Finally, you have to route IPv6 traffic over the tunnel interface. This is done by using the **ipv6 route** command, where the outgoing interface is set to the tunnel interface.

In the example, the CE1 router is assigned the 209.165.201.1 public IPv4 address on the GigabitEthernet0/0 interface. This IP address corresponds to the 2001:db8:d1a5:c901::/64 6to4 subnet. Therefore, the GigabitEthernet0/1 interface is configured with the IP address from the 2001:db8:d1a5:c901::/64 subnet.

Configuration of 6to4 relay routers is similar. The relay should be configured with the loopback interface that has the IPv4 address set to 192.88.99.1. This is a well-known IP address of 6to4 relays. This IP address should be advertised to the IPv4 network. On the relay, you have to configure the routing to route all traffic that is destined to 6to4 gateways to the tunnel interface. All other IPv6 traffic should be routed to the IPv6 Internet. In the example, traffic going to 2002:d1::/24 will be routed to the tunnel interface. The 2002:d1::/24 destination corresponds to all 6to4 traffic for 6to4 gateways that have the first octet of the IPv4 address set to 209 (that is, the first octet of the IP address of all CE routers in the example).



https://t.me/learningnets

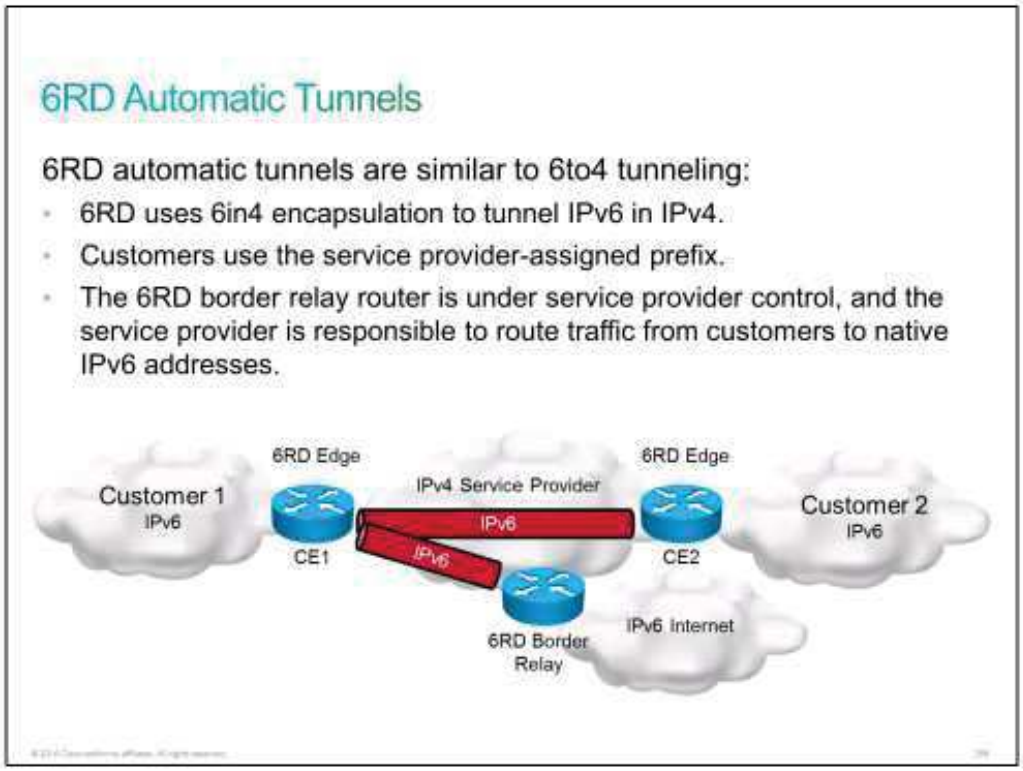
| | |
|---|---|
| **Note** | For a complete command reference, refer to the Cisco IOS and IOS XE Software Command Reference guides on www.cisco.com. |



https://t.me/learningnets

# 6RD Automatic Tunnels

6RD is an automatic tunnel establishment method that enables the connection of IPv6 islands through an IPv4 network. 6to4 uses 6in4 encapsulation to tunnel IPv6 in IPv4. In these aspects, 6RD is similar to 6to4.




The significant difference between 6RD and 6to4 is that 6RD operates entirely within the customer P-network and therefore avoids the drawbacks in the 6to4 design. Instead of using WKP for customer IPv6 addressing (as with 6to4), the service provider delegates one prefix from its addressing space. This prefix is then used to construct IPv6 addressing space for customers. Even if the service provider starts offering IPv6 access natively, addressing at the customer side can remain the same, and renumbering is not required.

The second difference between 6RD and 6to4 is that the 6to4 relay router, which is called the 6RD border relay, is under complete control of the service provider. The service provider is responsible to set up, operate, and maintain the 6RD border relay to enable customers to communicate with the IPv6 Internet.

 © 2014 Cisco Systems, Inc.

https://t.me/learningnets

## 6RD Automatic Tunnels (Cont.)

6RD addressing characteristics follow:

- The service provider selects a globally routable 6RD prefix from its IPv6 address space.
- The 6RD prefix is concatenated with the IPv4 address that is assigned to the customer. The resulting IPv6 network is assigned to the customer.
- Traffic between customers: The destination address falls within the 6RD prefix, and the tunnel destination is determined from the IPv6 prefix.
- Traffic to the IPv6 Internet: The destination address does not fall within the 6RD prefix, and traffic is sent to a preconfigured 6RD border relay.




A service provider selects a globally routable prefix from its address space and uses it as a 6RD prefix. In the example, the 6RD prefix is 2001:db8::/32, but it can also have a shorter or longer mask, if available.

Each 6RD site receives a prefix that concatenates the 6RD prefix (2001:db8::/32 in the example) and the IPv4 address of the 6RD CE edge router. For example, if the IPv4 address of the 6RD CE router is 209.165.201.1, the prefix of its IPv6 network will be 2001:db8:*IPv4 address of the edge router*::/64, which is 2001:db8:d1a5:c901::/64.

| | |
|---|---|
| **Note** | If the service provider uses the 6RD prefix with the /32 long mask and maps the entire customer IPv4 address, the network with mask /64 can be assigned to customers. The service provider can omit redundant parts of the IPv4 address (and the part that is common to all customers) and map only part of the IPv4 address. The service provider can either omit the prefix or suffix in the IPv4 address. Omitting the prefix or suffix depends on which part of the IP address is common to all customers. Thus, the service provider can assign a larger address space to customers (for example, /48 or /56). |

When an IPv6 packet goes to another IPv6 island, the destination address is in the 2001:db8::/32 range. When this packet reaches the 6RD CE router, the 6RD CE router extracts the IPv4 address that is embedded in the 2001:db8:: destination address. The 6RD CE router then will encapsulate the IPv6 packet in an IPv4 packet, with the IPv4 destination address extracted from the IPv6 destination address. This IPv4 address represents the address of the other 6RD CE router of the destination site. The destination 6RD CE router de-encapsulates the IPv6 packet in the IPv4 packet and then forwards the native packet toward its final destination.

When an IPv6 packet goes to the IPv6 Internet, the default route that points to the 6RD border relay will be used.

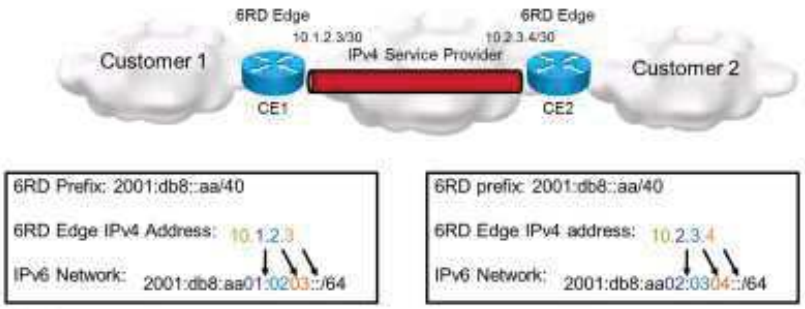


https://t.me/learningnets

## 6RD Automatic Tunnels (Cont.)

A 6RD addressing example follows:

- 6RD CE routers share the first octet of the IPv4 address.
- The service provider uses 2001:db8:aa::/40 as the 6RD prefix.
- The service provider would like to assign /64 networks to customers.



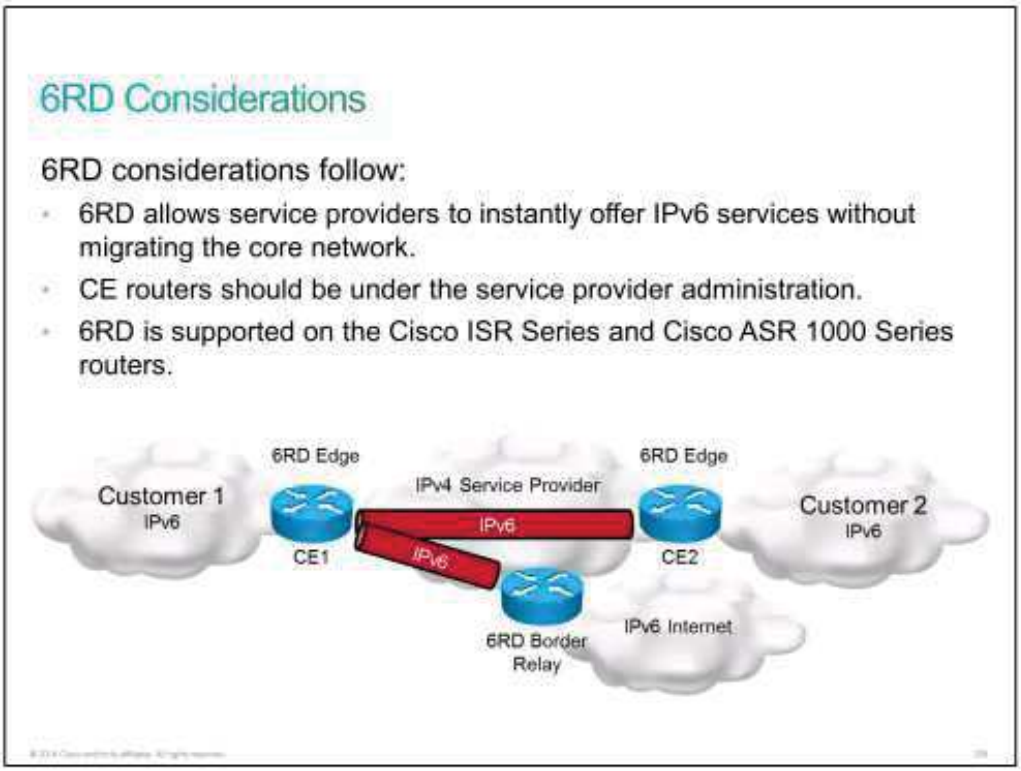

The figure shows an example of how to construct 6RD network addresses. All 6RD CE routers share the first octet of the IPv4 address. The service provider uses the IPv6 2001:db8:aa::/40 network as the 6RD prefix. The service provider would like to assign /64 networks to the customers. In order to assign these networks to the customers, the service provider has to omit the shared octet from the IPv4 address of the 6RD CE routers. The 40 bits of the 6RD prefix and 24 bits of the IPv4 address will be used to construct IPv6 networks. The IPv6 addresses are constructed by copying the 6RD prefix and then adding noncommon bits of the IPv4 address. The 6RD CE router with 10.1.2.3 IPv4 addresses therefore receives the 2001:db8:aa01:0203::/64 IPv6 network.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# 6RD Considerations

6RD tunneling is a recommended solution for service providers that want to instantaneously offer IPv6 service to customers without migrating the core network.




6RD tunneling solves the issues that arose with 6to4 tunnels; 6RD uses service provider-assigned IPv6 addresses and requires the service provider to maintain the 6RD border relay router.

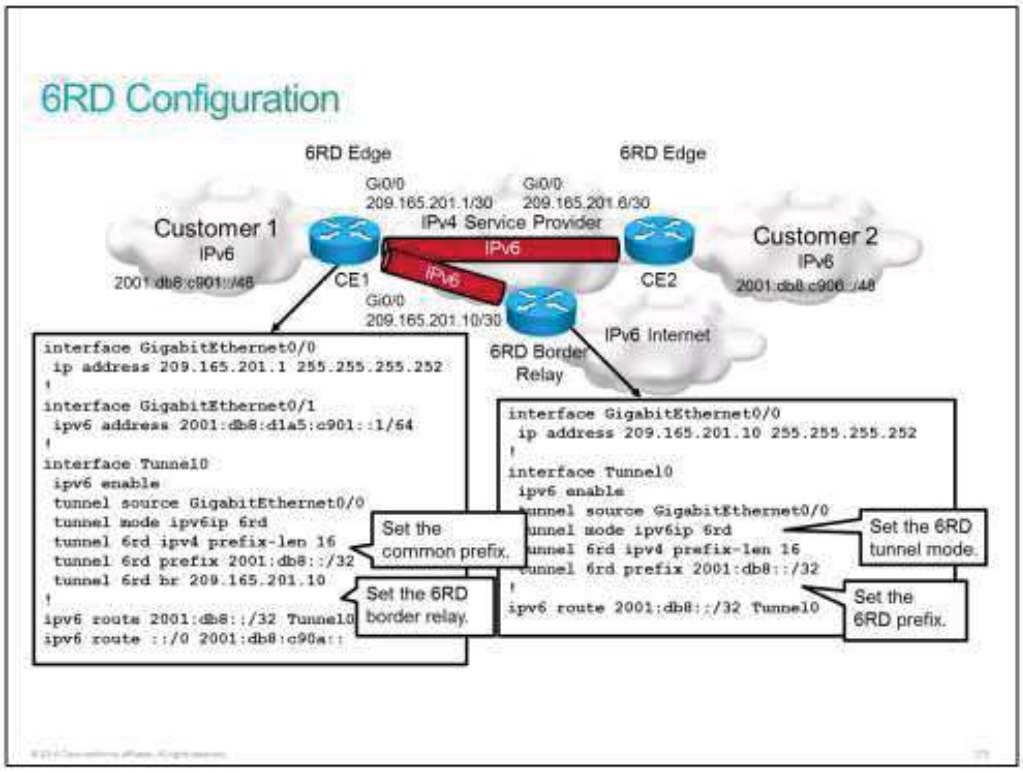
Because the 6RD CE routers require a 6RD configuration, it is recommended that CE routers be under the service provider administration.



https://t.me/learningnets

# 6RD Configuration

The 6RD tunnels are supported only on Cisco ISR Series routers and on Cisco ASR 1000 Series routers. Configuration for Cisco IOS and IOS XE Software is shown here:




To configure 6RD on the CE router, first create a tunnel interface using the **interface tunnel** command, followed by the interface identifier. Enable the tunnel interface for IPv6 using the **ipv6 enable** command. Then use the **tunnel source** command to specify the tunnel source, which is a physical interface that acts as a tunnel endpoint on the local router. Then set the tunnel mode to 6RD using the **ipv6ip 6rd** command. Specify the 6RD prefix using the **tunnel 6rd prefix** command. In the example, 2001:db8::/32 is used as the 6RD prefix. Then specify how many bits of the IPv4 address should be used to construct the IPv6 address. This is done using the **tunnel 6rd ipv4 prefix-len** command, followed by a number of bits that are common to all 6RD customers. In the example, the first 16 bits of the IPv4 address will not be used because the first 16 bits (209.165) are common to all customers. You can also specify the common suffix, if the customers share one, using the **suffix-len** keyword. Then specify the 6RD border relay IP address using the **tunnel 6rd br** command. In the example, the 6RD border relay is reachable at 209.165.201.10.

Finally, you have to route IPv6 traffic over the tunnel interface. Two separate routes are needed: one for traffic to other IPv6 islands, and one for the default route that points to the 6RD border relay. For traffic to other IPv6 islands, the route for the 6RD prefix should point to the tunnel interface. Then the default route should be configured that points to the next-hop IPv6 address, which is constructed using the 6RD prefix and noncommon part of the IPv4 address of the 6RD border relay. In the example, the next hop is set to 2001:db8:c90a:: because 201.10 corresponds to c90a.

Configuration of the tunnel interface on the border relay is similar, with the exception that the **tunnel 6rd br** command is not needed. The border relay should be configured with the IP address that is known to other 6RD routers as the 6RD border relay IP address. Finally, configure the border relay to route all traffic that is destined to 6RD routers (traffic going to the 6RD prefix) to the tunnel interface. All other IPv6 traffic should be routed to the IPv6 Internet.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

| | |
|---|---|
| **Note** | For a complete command reference, refer to the Cisco IOS and IOS XE Software Command Reference guides on www.cisco.com. |



https://t.me/learningnets

# Summary

This topic summarizes the key points that were discussed in this lesson.

## Summary

- Cisco IOS supports several ways of managing IPv4 address shortage; three of these are dual stack, CGN, and NAT64.
- Dual stack allows for coexistence of IPv6 and IPv4.
- Dual stack is less costly in the short term but more costly over the long term.
- To allow private IPv4 addresses to reach the Internet, you must implement CGN as well as dual stack.
- NAT444 means translating the IPv4 source address twice.
- Cisco supports carrier-grade NAT for deployment of the scalable NAT444 scenario.
- NAT64 means translating IPv6 addresses into IPv4 addresses and vice versa.
- NAT64 works together with (out of band) DNS64.

## Summary (Cont.)

- Stateless NAT64 can be used for static one-to-one translations.
- Stateful NAT64 is useful in general deployment.
- Stateless NAT64 is only useful in specific scenarios; for general purpose, use stateful NAT64.
- NAT64 is configured very similarly to regular NAT.
- For static NAT64, you must configure static translation in addition to NAT64.
- IPv6 tunneling is a convenient and simple way to send IPv6 across IPv4-only networks.
- IPv6 tunneling can be implemented in a manual or automatic way.
- GRE by definition allows tunneling of arbitrary Layer 3 protocols, and IPv6 is no exception.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

## Summary (Cont.)

- 6in4 is simpler method of tunneling but allows only IPv6 traffic in the tunnel.
- To set up 6in4 tunneling, you need to specify ipv6ip tunnel mode.
- 6to4 is an old method of building IPv6 tunnels that are specially suitable for scenarios where you do not have PA or PI address space.
- 6to4 tunneling should scale well and should be easy to configure due to automatic establishment of IPv6 tunnels.
- To create a 6to4 tunnel, the tunnel mode has to be set using the **ipv6ip 6to4** command.
- 6RD tunnels are an evolution of 6to4 tunneling that is specifically suited for rapid deployment within ISP networks.
- 6RD allows service providers to instantly offer IPv6 services without migrating the core network.
- To create a 6RD tunnel, the tunnel mode has to be set using the **ipv6ip 6rd** command.



https://t.me/learningnets

https://t.me/learningnets

# Lesson 3: Deploying IPv6 in the Service Provider Network

## Overview

There are several different ways to deliver IPv6 connectivity to customers. Service providers that are currently running only IPv4 in the P-network should understand deployment strategies. Understanding these strategies is important in order to offer IPv6 connectivity as soon as possible, without too much capital expenditure and disruption in existing IPv4 operations.

This lesson first describes the main methods that can be used by IPv4-only service providers to offer IPv6 connectivity to customers. These methods include dual stack, standalone IPv6 deployment on a separate Layer 2 infrastructure, and tunneling of IPv6 inside IPv4. The lesson then continues and concludes with popular last-mile technologies that allow customers to connect to IPv6-enabled P-networks.

Upon completing this lesson, you will be able to:

- Describe the main methods that IPv4-only service providers use to offer IPv6 services to their existing or new customers
- Describe the dual-stack option
- Discuss the advantages and disadvantages of the dual-stack option
- Describe the IPv6-in-IPv4 tunneling option
- Discuss the advantages and disadvantages of the IPv6-in-IPv4 tunneling option
- Discuss the two key strategies for phased service provider deployment of IPv6
- Describe the various IPv6 services
- Describe IPv6 address allocation
- Describe how service providers allocate IPv6 addresses to customers
- Describe broadband access services
- Describe the FTTH access architecture



https://t.me/learningnets

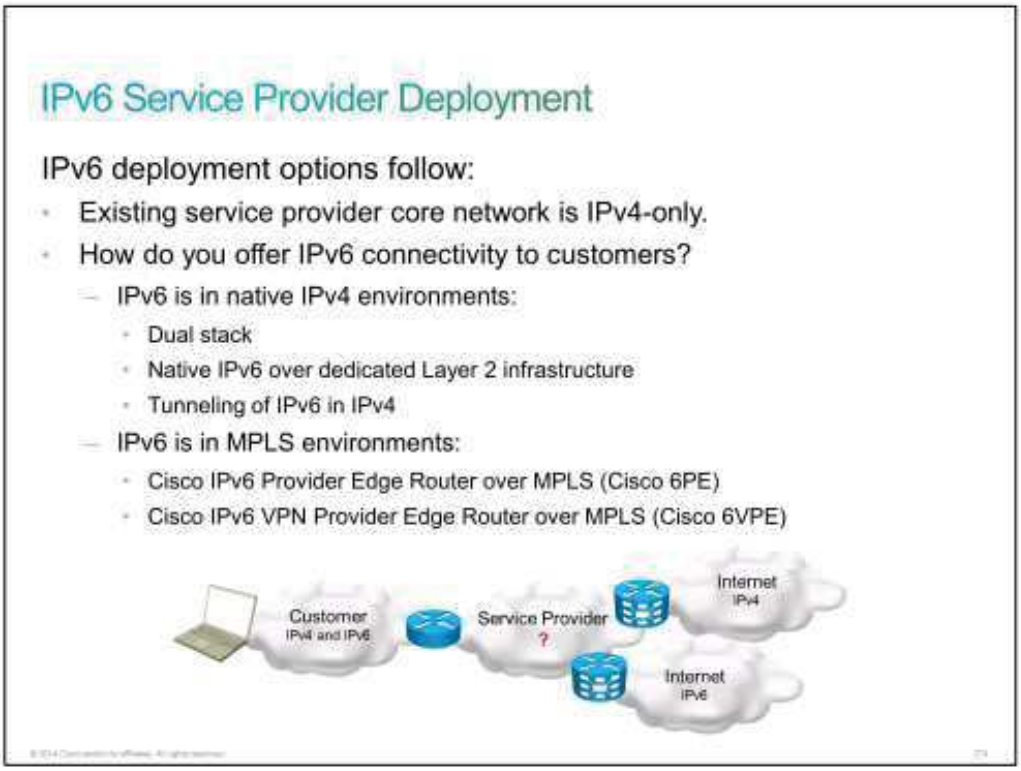
- Describe the DSL access architecture
- Describe the cable access architecture

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# IPv6 Service Provider Deployment

When the service provider runs IPv4 natively in the core network, there are three main methods that can be used to offer IPv6 services to existing or new customers. Each strategy has strengths and weaknesses that determine its ability to fit a particular service provider.




The three methods that can be used by IPv4-only service providers to offer IPv6 services to their customers follow:

- Upgrade the entire network to dual stack.
- Deploy a new parallel IPv6-only network on a dedicated Layer 2 infrastructure. Because this option is unlikely to be adopted by service providers, it is not discussed further.
- Upgrade the network edge and use tunneling of IPv6 through an IPv4-only core.

If the service provider runs IPv4 over MPLS, there are also other options to transfer IPv6 traffic over IPv4:

- **Cisco IPv6 Provider Edge Router over MPLS:** Cisco 6PE allows IPv6 domains to communicate with each other over an MPLS IPv4 core network. This implementation requires no core infrastructure upgrades and no reconfiguration of core routers because forwarding is based on the MPLS labels rather than on the IP header itself. This provides a cost-effective strategy for IPv6 deployment. For IPv6 transport to be transparent to all but Cisco 6PE routers, it is necessary to impose a hierarchy of labels at the Cisco 6PE ingress router. The top label provides connectivity inside the IPv4 MPLS core network: LDP or RSVP can be used to distribute the top label. The next label is used at each Cisco 6PE egress router for IPv6 forwarding and is distributed by MP-BGP.
- **Cisco IPv6 VPN Provider Edge Router over MPLS:** When IPv6 is enabled on an interface that is participating in an MPLS VPN, it becomes an IPv6 VPN. The CE to PE link is running IPv6 or IPv4 natively. The additional configuration of IPv6 on a PE router turns the PE into Cisco 6VPE, enabling service providers to support IPv6 VPNs over the MPLS network.



https://t.me/learningnets

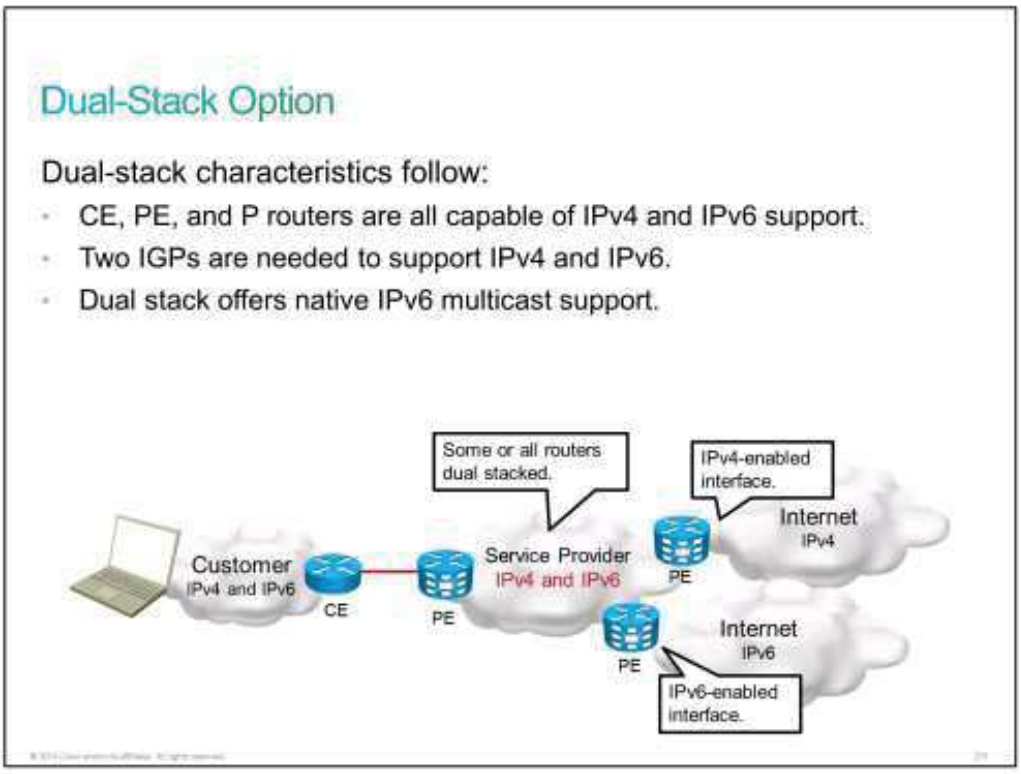
| | |
|---|---|
| **Note** | Cisco 6PE and 6VPE are outside of the scope of this course. These technologies are described in the *Implementing Cisco Service Provider Next-Generation Edge Network Services* (SPEDGE) v1.2 course. |

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# Dual-Stack Option

The dual-stack solution is clean and straightforward, providing customers with the most functionality and performance, but it can be a very large undertaking for the ISP. The solution involves upgrading all ISP network infrastructure, from the core to the edge to dual stack, where each link and interface is configured for both IPv4 (existing) and IPv6 (added). This process requires capital upgrades to network infrastructure in which nodes are not upgradable to IPv6 or do not have sufficient memory or other resources to run both protocols concurrently. The process also requires a change in routing policy and new peering arrangements. With this approach, however, customers receive the benefits of native IPv6 connectivity while maintaining the same IPv4 infrastructure.




The method of deploying IPv6 using dual-stack backbones allows IPv4 and IPv6 applications to coexist in a dual IP layer routing backbone. All routers in the network need to be upgraded to dual stack with IPv4 communication using the IPv4 protocol stack and IPv6 communication using the IPv6 stack.

Dual-stack end systems allow applications to migrate individually from IPv4 to IPv6 transport. Applications that are not upgraded that support only the IPv4 stack can coexist with upgraded applications on the same end system. Applications choose between IPv4 and IPv6 based on name lookups. Both the IPv4 and IPv6 addresses may be returned from the DNS, with the application selecting the correct address based on IP traffic type and particular communication requirements.

Rather than running different networks for different types of traffic (such as an IPv6-only network) or tunneling over the existing network (requiring network upgrades and re-engineering), dual stack allows native operation of IPv4 and IPv6.

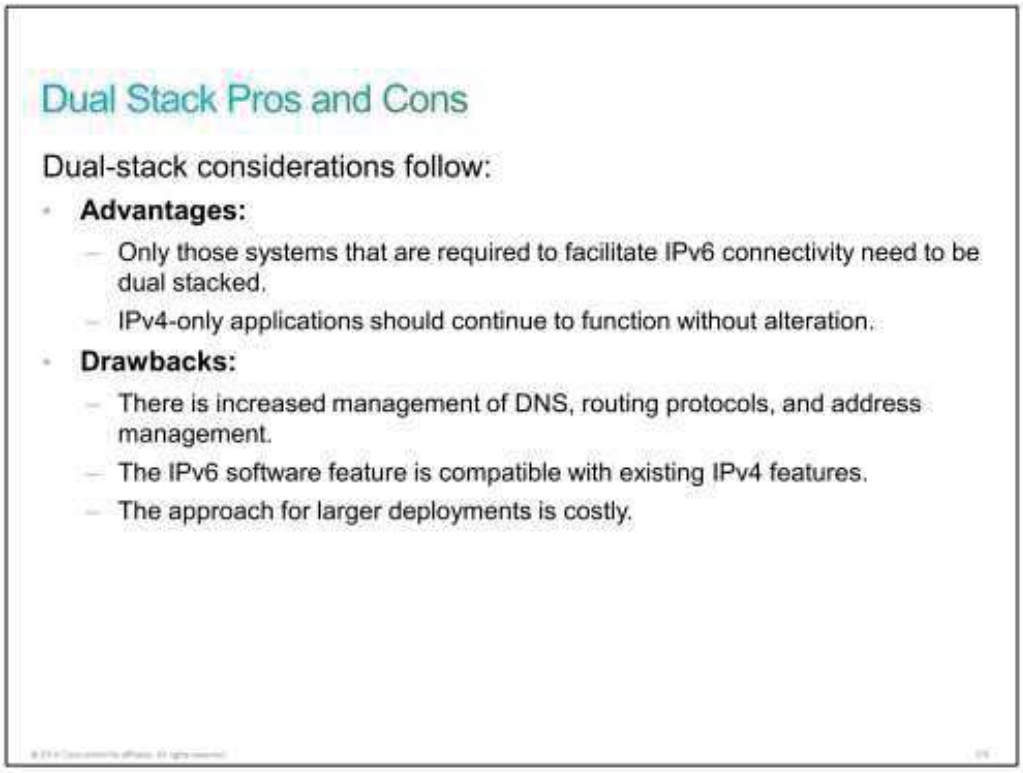
Dual stack introduces additional management overhead due to dual IP addressing and dual routing protocols. You also have to consider larger routing tables on the routers because of IPv4 and IPv6 routes. However, dual stack offers native support for IPv6 multicast.



https://t.me/learningnets

# Dual Stack Pros and Cons

Today, dual-stack routing is a valid deployment strategy for specific network infrastructures with a mixture of IPv4 and IPv6 applications (such as on a campus or an aggregation POP), requiring configuration of both protocols. However, apart from the obvious need to upgrade all routers in the network, limitations to this approach are that the routers require definition of a dual-addressing scheme, dual management of IPv4 and IPv6 routing protocols, and configuration with enough memory for both IPv4 and IPv6 routing tables.




The dual-stack approach has the following advantages:

- Only the devices that require IPv6 connectivity have to be dual stacked.
- IPv4 connectivity and applications should continue to work when IPv6 is deployed.
- Dual stack is easy to implement for small campus networks with a mixture of IPv4 and IPv6 applications.

Several drawbacks also must be considered:

- Dual stack may require dual management of routing protocols and a major upgrade for large networks.
- All routers must be dual stack with IPv4 and IPv6 addresses.
- Dual stack requires access to IPv6 DNS (AAAA records).
- Sufficient memory is required for both IPv4 and IPv6 routing tables. In practice, this has not been significant. The size of the new Cisco IOS Software images supporting advanced features and flash limitations of older equipment, can make this an issue. IPv6 software features may also introduce problems with existing IPv4 features.



https://t.me/learningnets

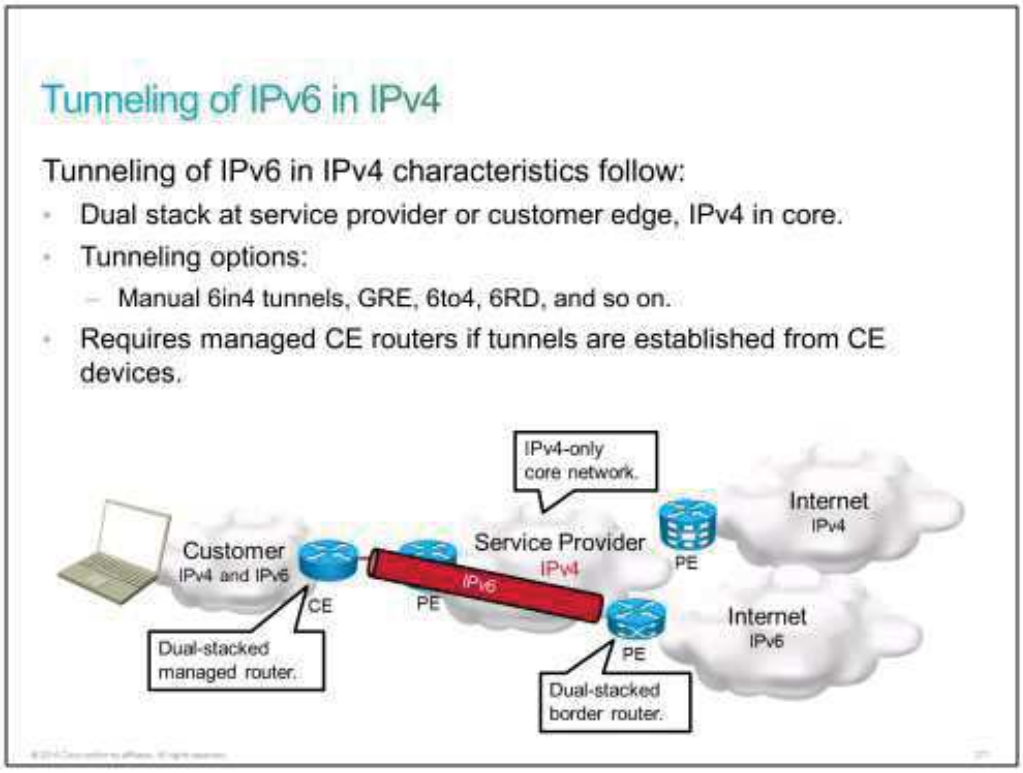
- The size and scope of dual stacking an entire enterprise network ensures that it will be costly in terms of people time and soft costs, if not in terms of capital investment as well. These costs are likely to greatly outweigh any equipment costs.



https://t.me/learningnets

# Tunneling of IPv6 in IPv4

The tunneling mechanism is used for communication between isolated IPv6 sites or connection to remote IPv6 networks over an IPv4 backbone. All tunneling mechanisms require that tunnel endpoints run both IPv4 and IPv6 (in dual-stack mode), while the core can remain IPv4-only. The dual-stack routers run both IPv4 and IPv6 protocols simultaneously and therefore can interoperate directly with both IPv4 and IPv6 end systems and routers. In many senses, tunneling IPv6 in IPv4 is a hybrid strategy that still requires dual stack at tunnel endpoints.




As mentioned previously and discussed with details in the previous lesson, there are several tunneling methods that are available to enable communication between isolated IPv6 sites. The options are to use manual 6in4 encapsulation tunnels, GRE tunnels, 6to4 tunnels, and 6RD tunnels, with the latter as the recommended option. Tunneling requires managed CE routers if tunnels are established from CE routers.

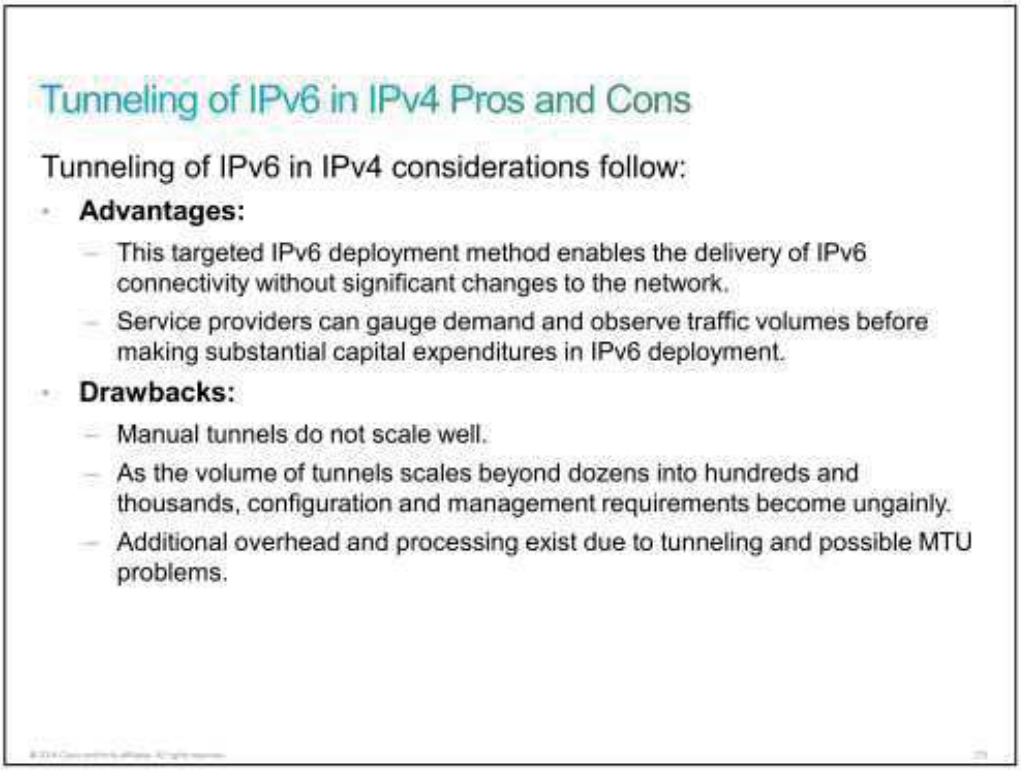
The figure shows a dual-stack customer environment, which uses tunneling through an existing IPv4 P-network to dual stack a border router that transits customer traffic to the IPv6 Internet.



https://t.me/learningnets

# Tunneling of IPv6 in IPv4 Pros and Cons

The advantages and disadvantages of the IPv6 in IPv4 tunneling option are described here:




The advantages for a service provider of using configured IPv6 tunnels to provide IPv6 services to customers include the following:

- Provides IPv6 connectivity to IPv6 customer sites without disrupting the existing IPv4 service of the service provider and without a large financial or resource commitment from the ISP. Creates new revenue streams for the service provider by offering new IPv6 services and support for new IPv6 applications.
- Enables the service provider to validate and gauge future demand for IPv6 while generating revenue from initial customer sites that request IPv6 services. Enables the service provider to monitor individual traffic statistics for the customer site and charge the customer site for IPv6 services.

The disadvantages of the tunneling approach are as follows:

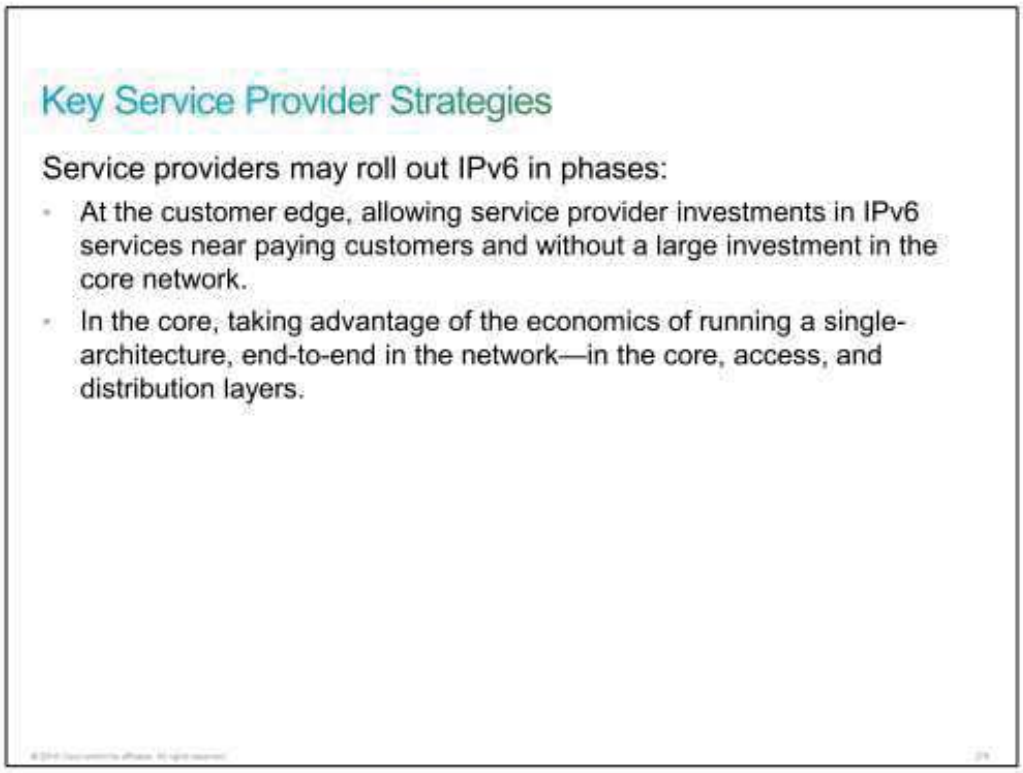
- A service provider using manually configured IPv6 tunnels to provide IPv6 services to customers is supporting additional IPv6 customers, which requires the service provider to manually configure additional IPv6 tunnels. Management overhead increases as the number of tunnels increases.
- At some point, the number of manually configured IPv6 tunnels that each ISP border router can support might be exhausted, and the ISP will need to upgrade the border router hardware. The ISP then will need to re-engineer its IPv6 service offering.
- Tunneling introduces additional overhead and processing on tunnel endpoints and lowers the MTU that can be used to transfer data.



https://t.me/learningnets

# Key Service Provider Strategies

The strategies for phased service provider deployment of IPv6 are described here:




There are two key strategies for phased service provider deployment of IPv6:

- **Providing an IPv6 service at the customer access level:** Starting the deployment of IPv6 at the customer access level permits an IPv6 service to be offered immediately without a major upgrade to the core infrastructure and without affecting current IPv4 services. This approach allows an evaluation of the IPv6 products and services before full network implementation as well as an assessment of the future demand for IPv6 without substantial investment early in the process.
- **Running IPv6 within the core infrastructure itself:** At the end of the initial evaluation and assessment stage, as support for IPv6 within the routers improves (particularly IPv6 high-speed forwarding) and as network management systems fully embrace IPv6, the network infrastructure can be upgraded to support IPv6. This upgrade path could involve dual-stack routers or, as IPv6 traffic gradually becomes predominant, IPv6-only routers.

Service providers that are ready to begin the IPv6 deployment process should consider these issues:

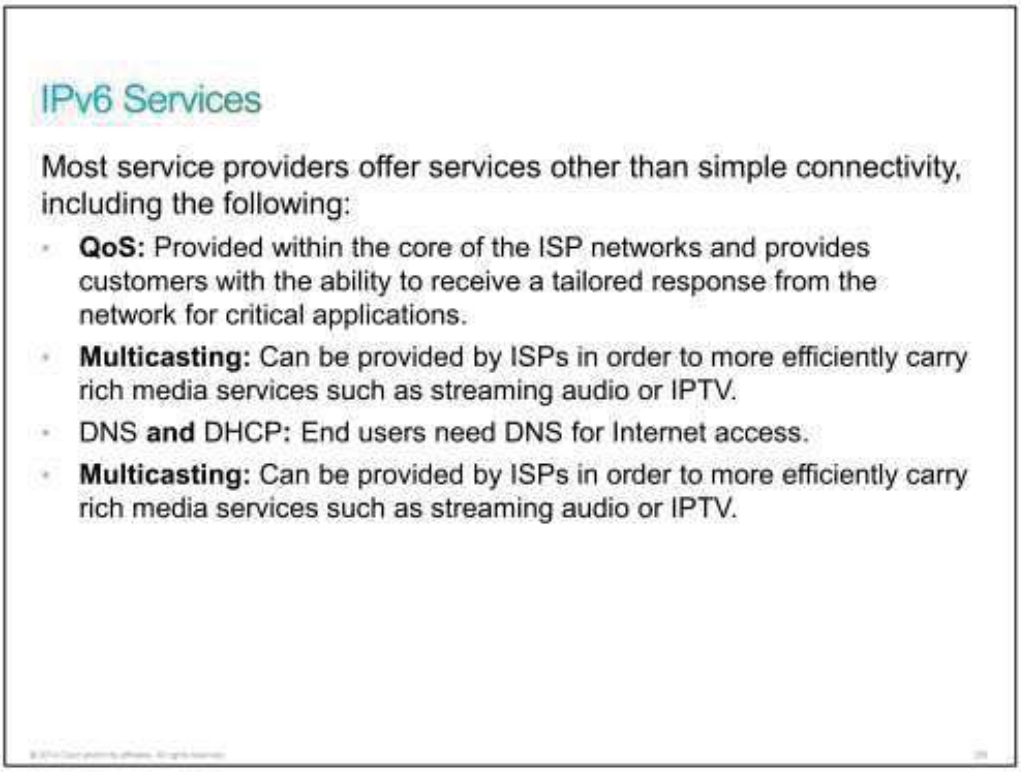
- Service providers need IPv6 global addresses (delegation)
- Migration of the DNS infrastructure
- Network migration
- Customers access control and accounting



https://t.me/learningnets

# IPv6 Services

Most service providers offer services beyond basic connectivity, Internet access, and IP transport. For smaller service providers that are closer to the end consumer, products such as web hosting and email are quite common. Larger ISPs may also provide dedicated circuits and features such as QoS support and IP multicasting. In many cases, service provider services will be affected by the integration and support of IPv6.




Some of the services that are affected by IPv6 integration are as follows:

- **QoS support:** The introduction of IPv6 should not impair existing SLAs or similar quality reassurances. Depending on the deployment of the IPv6 service, the service could be best effort, at least initially, even if the IPv4 service had an SLA. Both IntServ and DiffServ are equally applicable in IPv6 and IPv4, working in similar fashion. Of these two, typically only DiffServ is implemented.

  IPv6 and IPv4 both currently have the same QoS capabilities. IPv6 has area service differentiation as an add-on that is derived from an IPv6 flow label field. In the future, this field could add pricing information along with service differentiation. QoS is also increased indirectly. Always-on connections, increased network performance, and similar features also increase QoS. As a network improves, customer expectations also rise. Better-quality networks alone probably will not increase customer satisfaction. A service provider offering these connectivity services should not forget that competitors are improving their networks as well.

  In the future, charging for access quality will likely become more common as many networks implement QoS. This charging model makes possible a guaranteed, predetermined level of network capacity between specific locations that is crucial for certain clients. This model also makes possible the dynamic provision of different service levels to a customer, which is based on what the customer is willing to pay for access at any given time.



https://t.me/learningnets

QoS features that are supported for IPv6 environments include packet classification, queuing, traffic shaping, WRED, class-based packet marking, and policing of IPv6 packets. These features are available at both the process switching and Cisco Express Forwarding switching paths of IPv6.

- **Multicasting services:** IPv6 multicasting is well developed, both on the network and host sides, as implementations exist for Windows XP, Linux, Solaris, and so on. Many service provider implementations of multicast on IPv4 use tunnels because few service providers have fully multicast-enabled networks. This scenario can result in an implementation in which multicast traffic is tunneled IPv6 in IPv6 for IPv6 networks or in which multicast traffic is tunneled IPv6 in IPv4 between multicast-enabled routers. Both tunneling and multicast have become well-understood concepts.

  Multicast can operate simply using the unicast routing table as a base to build the multicast routing information, but most service providers run a multicast-enabled routing protocol in order to manage the networks separately. MP-BGP is used as the interdomain protocol.

- **DNS and DHCP:** DNS is required for normal operation of Internet clients and must be offered by a service provider. If clients use stateless autoconfiguration, their DNS settings must either be static or acquired over DHCP. In the latter case, a service provider also must provide DHCP service.

- **Managed services:** Whenever the client lacks resources or technical knowledge to set up and maintain a set of services, the service provider may offer to do this for the client.

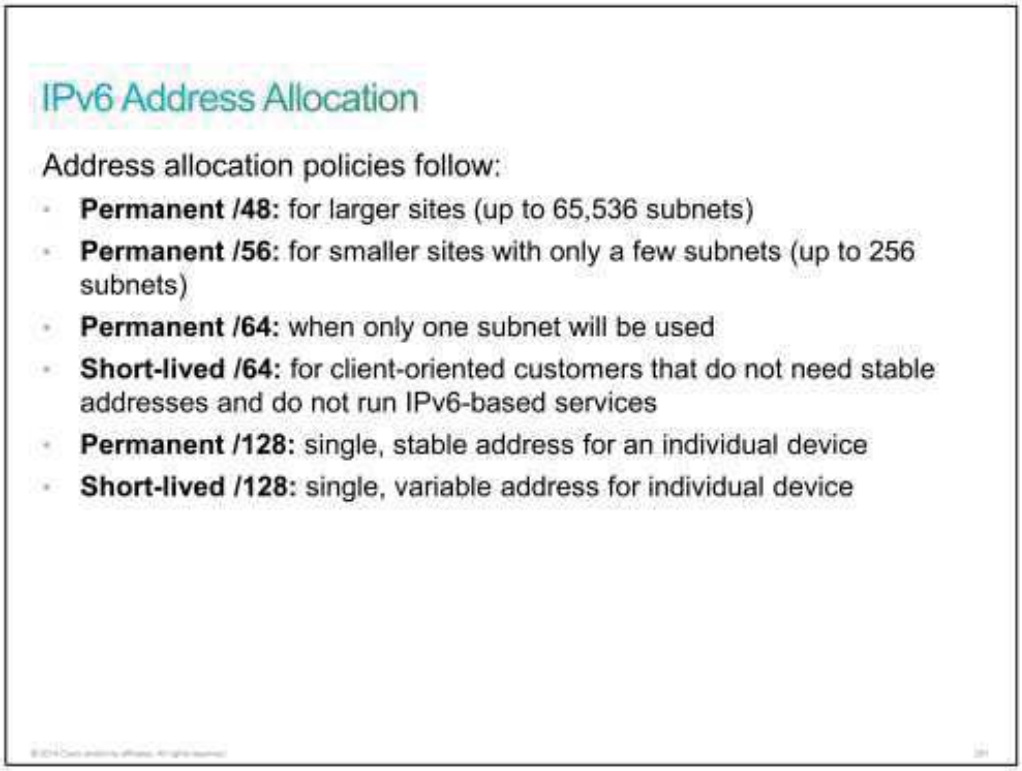
- **Managed security:** Like managed services, if a service provider offers managed security, it offloads security checking, policy maintenance, and equipment from the customer to itself. The customers need not set up and maintain security systems; the service provider does it for them. In some cases, this may be unacceptable for the customers, if they wish to have complete control over all communications. Offloading some services to the service provider may put the service provider in a position where critical systems are not under customer control.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# IPv6 Address Allocation

The principle for address allocation is stated as follows: "…recommends the assignment of /48 in the general case, /64 when it is known that one and only one subnet is needed…" —RFC 3177, *IAB/IESG Recommendations on IPv6 Address Allocations to Sites*.




Broadband address allocation for IPv6 presents several choices for the service provider. Several different addressing schemes are available, depending on the customer environment, the need for permanent or temporary addresses, and the need to support single hosts, small networks, or large networks.

The most common implementations follow:

- **Permanent /48:** This allocation provides a customer the same /48 each time they connect. The /48 prefix allocation means that the customer can have multiple networks that are all capable of doing autoconfiguration (which requires the /64 prefix).
- **Permanent /56:** This allocation is similar to a permanent /48, except for a lower number of possible subnets at the customer site.
- **Permanent /64:** This allocation supports a customer site that has either a single device or a router with a single stub network behind it that desires stable addresses and autoconfiguration.
- **Short-lived /64:** This allocation supports the same environment as a permanent /64 but does not provide the same /64 network to the customer each time they connect.
- **Permanent /128:** This static allocation is used to connect single devices because only a single IPv6 address is allocated to the customer.
- **Short-lived /128:** This variable allocation is used to connect single devices because only a single IPv6 address is allocated to the customer.



https://t.me/learningnets

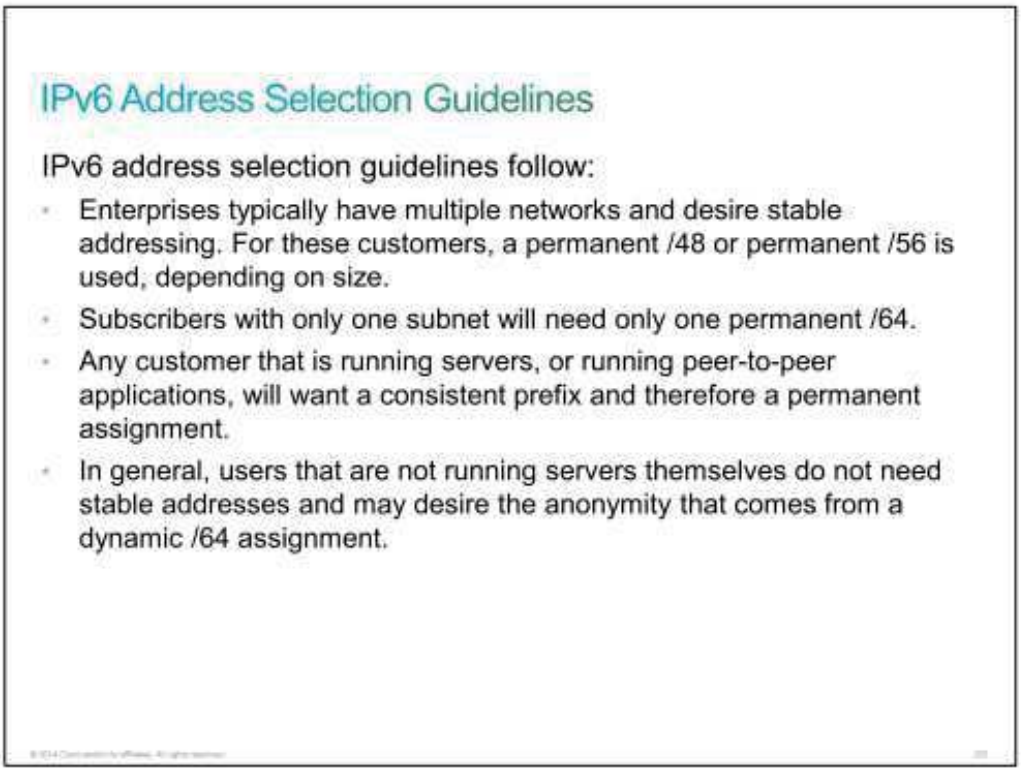
Address allocation is independent of the method that is used by the customer to connect to the NAP. These alternatives are determined for each customer ("per-user"), and configuration information is set up on the NAS and on a RADIUS server that is used both for customer authentication and to establish connection parameters.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# IPv6 Address Selection Guidelines

Service providers will use various methods for allocating addresses to customers, and most service provider implementations will have several in use at all times, all of which are driven by the information that is stored in a RADIUS server (or an alternative AAA solution).




The following guidelines can be used to determine which solution is best for a given situation:

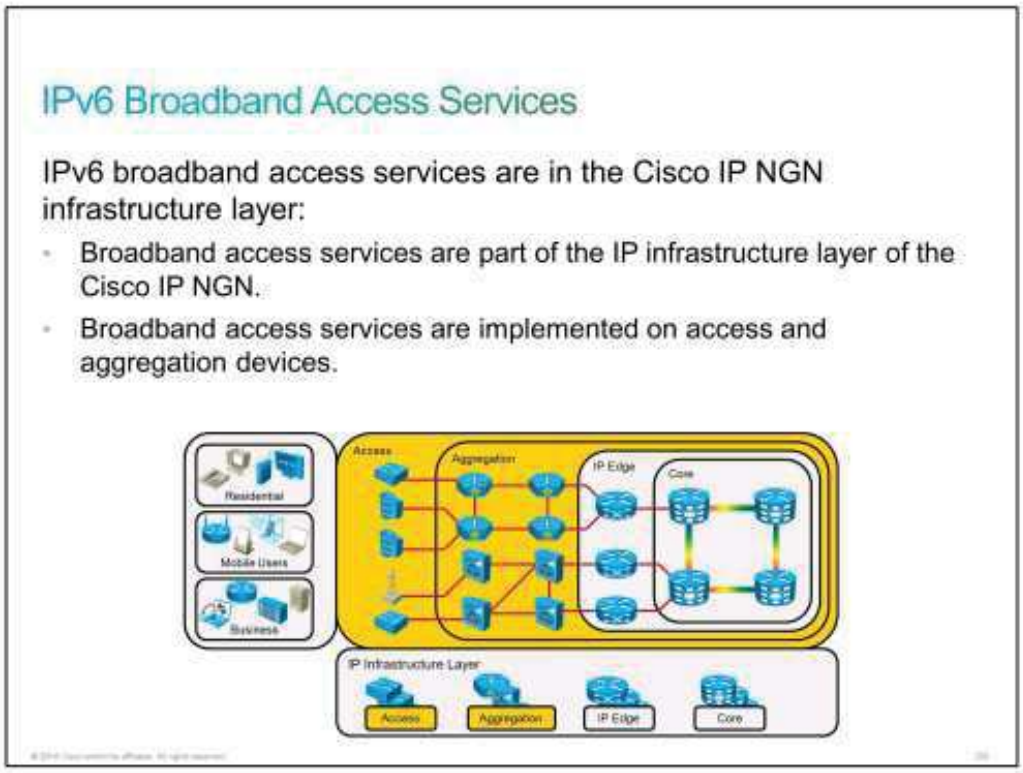
- Understand the needs of the customers in terms of how many devices and networks that they need to support.
  - If the customer has multiple networks, the recommended allocation is /48, which gives them enough subnet bits for 64K (128–48–64 bits = 16 bits, 2^16 = 64K) networks, in which each network will support autoconfiguration. Smaller allocations are also technically possible, such as /60 (which would allow for 16 [128–60–64 bits = 4 bits, 2^4] customer networks). These allocations are always permanent; to make them temporary, would require ongoing renumbering of the enterprise network.
  - If the customer has only a single link or a single device (host), the recommended allocation is /64. This allows for the CPE, when it is a router, to use the bridging mode to build a multilink network. In the network, hosts behind the router can autoconfigure addresses from the same /64 that is used to connect the CPE to the PE.
- For customers needing a /64 or a /128, determine if the allocation should be permanent:
  - If the customer will run peer-to-peer or server IPv6 applications or act as a server, the address assignment should be permanent so that stable addresses can be put into DNS. The customer will always obtain the same prefix for a /64, where the customer can use static IPv6 address assignment.
  - For customers that are using client-only applications or desiring a measure of privacy from a changing prefix or address, transient allocation is acceptable and may be preferred.



https://t.me/learningnets

# IPv6 Broadband Access Services

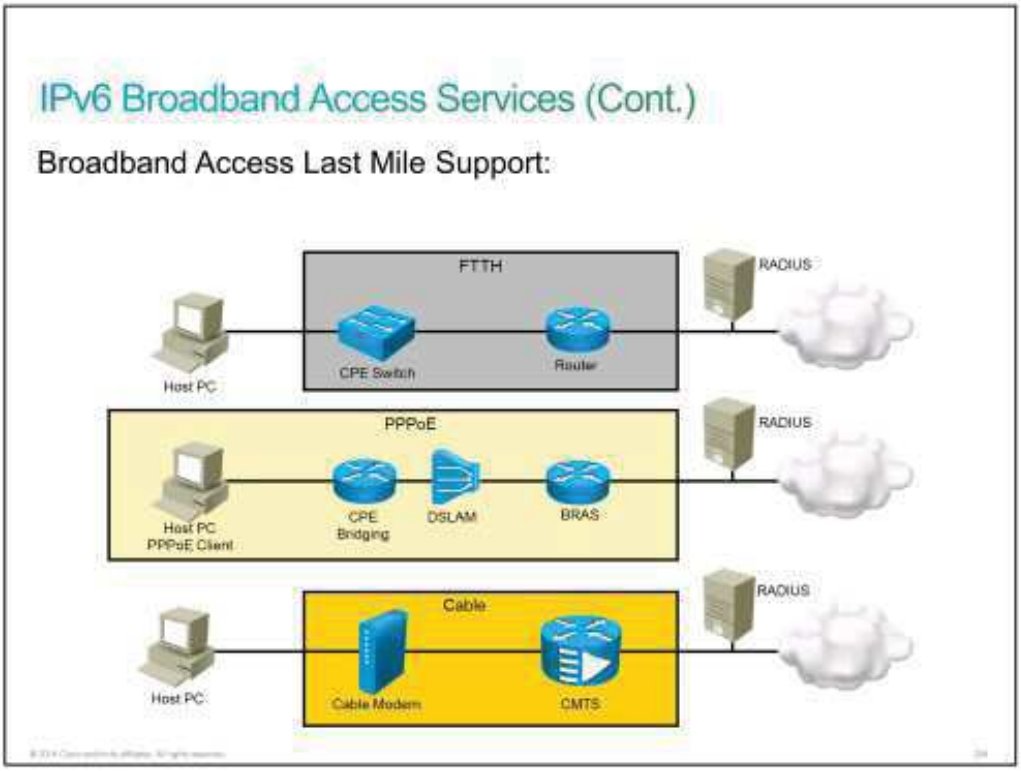
The IPv6 broadband access services are described here:




Broadband access services are implemented on access and aggregation devices of the Cisco IP Next-Generation Network framework. Broadband access services enable end users and customers to connect to the P-network over the "last mile."



https://t.me/learningnets




To benefit from the huge addressing space of IPv6 and bring IPv6 broadband access to end customers, last-mile access has to support IPv6 as well. Broadband architectures for Internet access among others offer the following most-widespread, last-mile implementations:

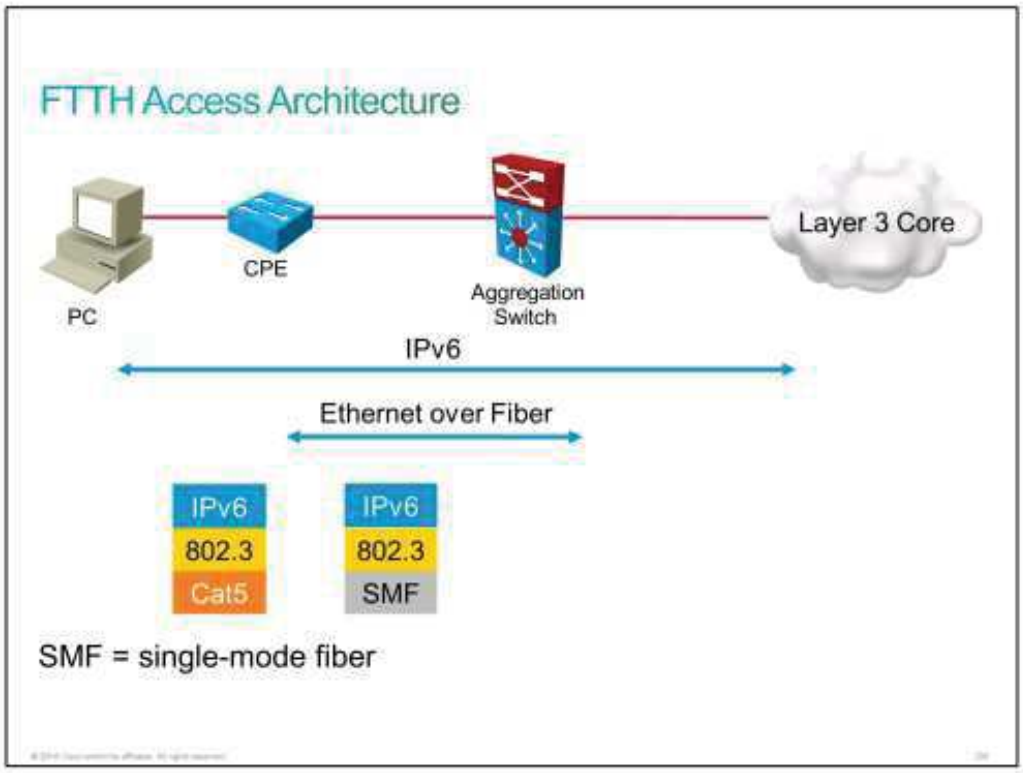
- FTTH: FTTH gained popularity recently with the advent of fiber lines all of the way to the residential buildings. The CPE device typically is a switch that has optical uplink and copper downlinks, where end users connect their own equipment or end nodes directly.
- PPPoE: PPPoE architecture typically is a CPE router with Ethernet on the customer side and ATM on the ISP-facing side. The CPE router can initiate the PPP session to the service provider PE or NAS, or the PPP session can be originated from each PC within the Ethernet network. The backend infrastructure is the same as for PPPoA.
- **Cable network:** Cable networks reach all homes with cable TV. In these networks, the CPE device is a cable modem. Modern modems may have router capabilities; otherwise, they bridge Ethernet segments over the CATV network.



https://t.me/learningnets

# FTTH Access Architecture

The figure displays typical components in FTTH access architecture.



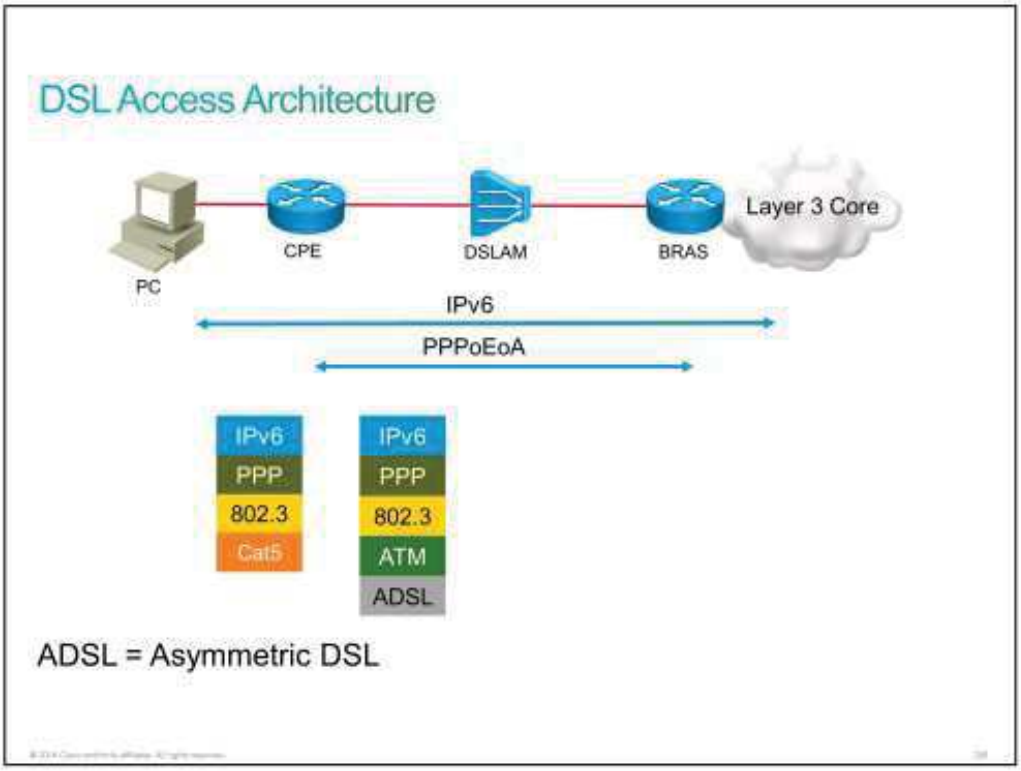

End hosts are directly connected to the P-network via Ethernet switches. There is no encapsulation if the service provider supports native IPv6. IPv6 packets are encapsulated inside Ethernet just like IPv4 packets, with Ethertype set to 0x86DD.



https://t.me/learningnets

# DSL Access Architecture

The figure displays typical components in DSL access architecture.




The transport protocol that is used in the DSL network is ATM. The DSL modem encapsulates the DSL customer PPP packets inside ATM cells and sends them over the WAN.

End hosts are directly connected to the P-network via CPE devices, and PPP sessions are terminated on the IPv6-enabled BRAS. Between the end hosts (or router, if PPP is terminated on a router) and CPE device, the PPP session is encapsulated inside Ethernet. Between CPE and BRAS, the PPP session is usually encapsulated inside Ethernet, which is bridged over ATM.

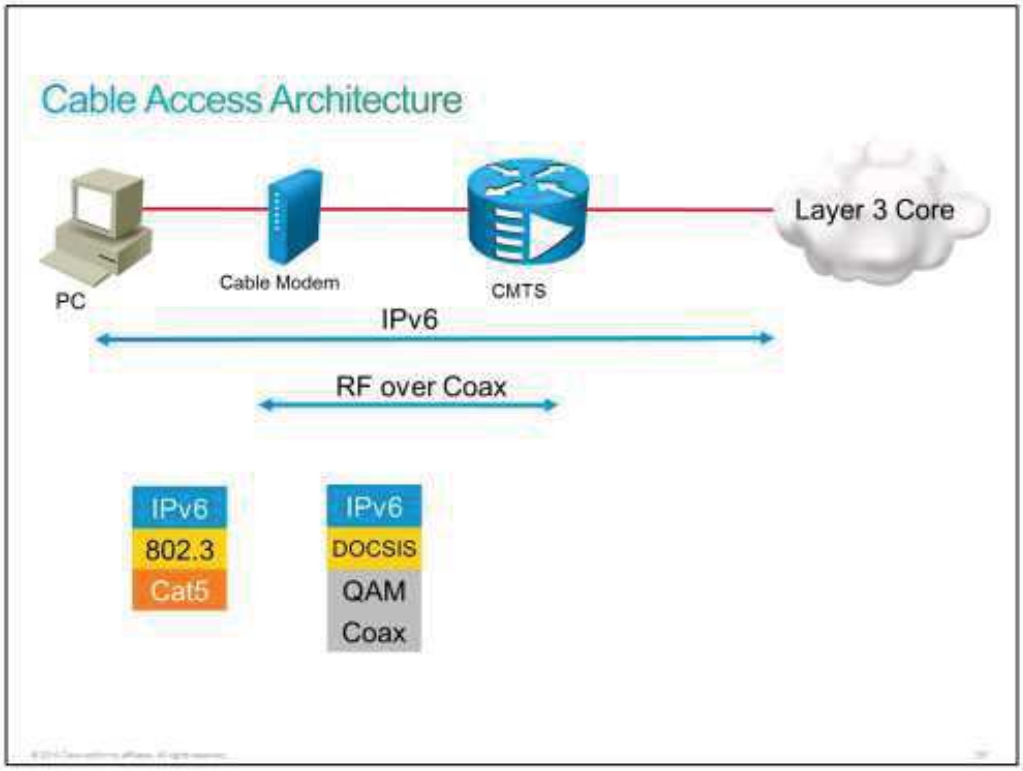
PPP in general has two phases. In the first one, called LCP, the PPP device sends LCP packets to configure and test the data link. LCP packets contain a configuration option that allows devices to negotiate the use of options, such as the maximum receive unit, compression, and the link authentication protocol. In the second phase, called NCP, PPP chooses and configures the network layer protocol, such as IPv4 or IPv6. NCP for IPv4 is called IPCP, and NCP for IPv6 is called IPv6CP.



https://t.me/learningnets

# Cable Access Architecture

The figure displays typical components in cable network access architecture.




End hosts are connected to the P-network via coaxial cable network. CPE is a modem, which is connected to a CMTS over cable network. Ethernet is bridged across, and IPv6 can be encapsulated in Ethernet such as IPv4. IPv6 packets are transferred over the coaxial cable inside DOCSIS frames, which are modulated using a kind of QAM. Support for IPv6 was introduced in DOCSIS version 3.0.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# Summary

This topic summarizes the key points that were discussed in this lesson.

## Summary

- Service providers that only have IPv4 networks can use workarounds to still provide IPv6 to end users while maintaining the IPv4-only core network.
- Implementing IPv6 side by side with IPv4 will not disrupt IPv4, but it may increase network core complexity.
- There are many advantages when deploying dual stack, but the drawbacks should be considered.
- Implementing IPv6 tunneling will increase complexity only at the edge and may also affect scalability depending on the type of tunneling.
- Tunneling of IPv6 in IPv4 enables the delivery of IPv6 connectivity without significant changes to the network, but additional overhead and processing exist.
- Service providers will want to invest step by step into new technologies at the edge and avoid large investments into the core by keeping changes away from it.

## Summary (Cont.)

- Network services that may be affected by the introduction of IPv6 are QoS, multicast, DNS, DHCP, and managed services.
- Assignment of IPv6 addresses is different than assignment of IPv4 addresses because IPv6 addresses are not as scarce. Policies must therefore be different to allow customers to take advantage of all IPv6 features.
- ISPs should assign address space based on the number of segments that are required and not the number of nodes that are present.
- Most Layer 2 technologies support IPv6, and you can easily deploy IPv6 over the following:
  - FTTH access architecture
  - DSL access architecture
  - Cable access architecture



https://t.me/learningnets

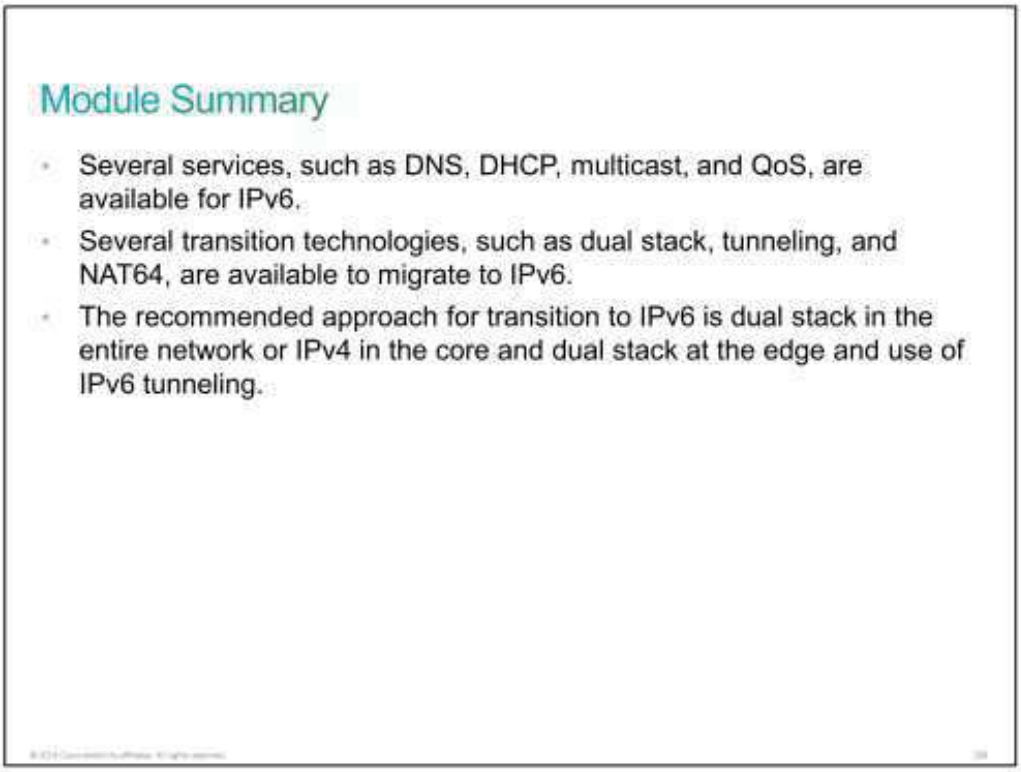
 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# Lesson 4: Module Summary

## Overview

This topic summarizes the key points that were discussed in this module.




This module first described services that are available on IPv6. These included DNS, DHCPv6, IPv6 multicast, and IPv6 QoS. The module also discussed transition technologies that are available to migrate to IPv6. The module concluded with deployment strategies that service providers are facing when deploying IPv6.



https://t.me/learningnets

https://t.me/learningnets

# Lesson 5: Module Self-Check

Use the questions here to review what you learned in this module. The correct answers and solutions are found in the Module Self-Check Answer Key.

1. Which two IPv6 multicast addresses are valid and can be used as group addresses to stream multimedia content in a service provider network? (Choose two.) (Source: Introducing IPv6 Services)

   A. ff02::5
   B. ff1e::100:100
   C. ff18::abcd
   D. fe18::abcd

2. The Cisco switch performing MLD snooping forwards only the first report for a multicast group to the router and suppresses all other reports for the same multicast group. (Source: Introducing IPv6 Services)

   A. True
   B. False

3. Which two DNS record types are valid? (Choose two.) (Source: Introducing IPv6 Services)

   A. A6
   B. AAAA
   C. MX
   D. PTR6

4. A DHCPv6 client has to examine router advertisements before requesting an IPv6 address. (Source: Introducing IPv6 Services)

   A. True
   B. False



https://t.me/learningnets

5. Which two fields can be used to match IPv6 packets for QoS? (Choose two.) (Source: Introducing IPv6 Services)

   A. ToS
   B. Traffic class
   C. Source IP address
   D. Flow label

6. Which management protocol or application is not available for IPv6 on the Cisco IOS XR Software platform? (Source: Introducing IPv6 Services)

   A. SNMP
   B. Syslog
   C. IP SLA
   D. NTP

7. Which two technologies can be used to offer services running on IPv4 servers to IPv6-only hosts? (Choose two.) (Source: Defining IPv6 Transition Mechanisms)

   A. Carrier-grade NAT
   B. Static stateful NAT64 with IPv6-enabled DNS
   C. Stateless NAT64 with DNS64
   D. 6to4 tunnels

8. Which option presents a valid 6to4 IPv6 prefix for the 192.168.101.15 address? (Source: Defining IPv6 Transition Mechanisms)

   A. 2001:c0a8:650f::/48
   B. 2002:c0a8:65f0::/48
   C. 2001:db8:c0a8:65f0/64
   D. 2002:c0a8:650f::/48

9. 6RD is a recommended method for tunneling IPv6 traffic across the IPv4 core. (Source: Deploying IPv6 in the Service Provider Network)

   A. True
   B. False

10. Match the following last-mile access technologies with a standard supporting IPv6.(Source: Deploying IPv6 in the Service Provider Network)

| | |
|---|---|
| PPPoE | Ethertype 0x86DD |
| cable | IPv6CP |
| FTTH | DOCSIS 3.0 |

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

# Module Self-Check Answers

## Answer Key

1. B, C
2. A
3. B, C
4. A
5. B, C
6. C
7. B, C
8. D
9. A
10. FTTH — Ethertype 0x86DD
    PPPoE — IPv6CP
    cable — DOCSIS 3.0



https://t.me/learningnets

https://t.me/learningnets

# Glossary

**6in4**

6in4 is a transition mechanism for migrating from IPv4 to IPv6. 6in4 uses tunneling to encapsulate IPv6 traffic over explicitly-configured IPv4 links as defined in RFC 4213. The 6in4 traffic is sent over the IPv4 Internet inside IPv4 packets whose IP headers have the IP protocol number set to 41.

**6RD**

IPv6 rapid deployment. 6RD is a mechanism to facilitate IPv6 rapid deployment across IPv4 infrastructures of Internet service providers (ISPs).

**6to4**

6to4 is a transition mechanism for migrating from IPv4 to IPv6. 6to4 allows IPv6 packets to be transmitted over an IPv4 network without the need to configure explicit tunnels. The method 6to4 makes use of protocol 41, instead of static configuration of the endpoints, the endpoint IPv4 address information is derived from the IPv6 addresses within the IPv6 packet header.

**AAA**

authentication, authorization, and accounting. Pronounced "triple a."

**AAAA**

quad A

**ARP**

Address Resolution Protocol. Internet protocol that is used to map an IP address to a MAC address. Defined in RFC 826.

**AS**

autonomous system. A collection of networks under a common administration sharing a common routing strategy. Autonomous systems are subdivided by areas. An autonomous system must be assigned a unique 16-bit number by the IANA.

**ATM**

Asynchronous Transfer Mode. The international standard for cell relay in which multiple service types (such as voice, video, or data) are conveyed in fixed-length (53-byte) cells. Fixed-length cells allow cell processing to occur in hardware, thereby reducing transit delays. ATM is designed to take advantage of high-speed transmission media, such as E3, SONET, and T3.

**Auto-RP**

Auto-Rendezvous Point

**BGP**

Border Gateway Protocol. Interdomain routing protocol that replaces EGP. BGP exchanges reachability information with other BGP systems. It is defined by RFC 1163.

**BIDIR-PIM**

Bidirectional PIM



https://t.me/learningnets

**BRAS**

Broadband Remote Access Server

**BSR**

Bootstrap Router

**CATV**

cable television. A communication system where multiple channels of programming material are transmitted to homes using broadband coaxial cable. Formerly called Community Antenna Television.

**CBWFQ**

class-based weighted fair queuing. Extends the standard WFQ functionality to provide support for user-defined traffic classes.

**ccTLD**

country code top-level domain

**CE**

customer edge. Identifies the network devices, connected to a provider network, that are under the administrative control of the customer.

**CGN**

carrier-grade NAT

**CMTS**

cable modem termination system

**CoS**

class of service. An indication of how an upper-layer protocol requires a lower-layer protocol to treat its messages. In SNA subarea routing, CoS definitions are used by subarea nodes to determine the optimal route to establish a given session. A CoS definition comprises a virtual route number and a transmission priority field. Also called ToS.

**CPE**

customer premises equipment. Terminating equipment, such as terminals, telephones, and modems, supplied by the telephone company, installed at customer sites, and connected to the telephone company network. Can also refer to any telephone equipment residing on the customer site.

**DF**

designated forwarder

**DHCP**

Dynamic Host Configuration Protocol. Provides a mechanism for allocating IP addresses dynamically so that addresses can be reused when hosts no longer need them.

**DHCPv6**

Dynamic Host Configuration Protocol version 6



https://t.me/learningnets

**DiffServ**

differentiated services. A paradigm for providing QoS on the Internet by employing a small, well-defined set of building blocks from which a variety of services can be built.

**DNS**

Domain Name System. System used on the Internet for translating names of network nodes into addresses.

**DNS64**

Domain Name System 64. DNS64 describes a DNS server that when asked for a domain's AAAA records, but only finds A records, synthesizes the AAAA records from the A records.

**DOCSIS**

Data-over-Cable Service Interface Specifications. Defines technical specifications for equipment at both subscriber locations and cable operator headends.

**DR**

designated router.

**DSCP**

differentiated services code point.

**DSL**

digital subscriber line. Public network technology that delivers high bandwidth over conventional copper wiring at limited distances. There are four types of DSL: ADSL, HDSL, SDSL, and VDSL. All are provisioned via modem pairs, with one modem located at a central office and the other at the customer site. Because most DSL technologies do not use the whole bandwidth of the twisted pair, there is room remaining for a voice channel.

**EF**

Expedited Forwarding. A differentiated services per-hop behavior intended to provide a guaranteed bandwidth with low delay, jitter, and loss.

**EIGRP**

Enhanced Interior Gateway Routing Protocol. Advanced version of IGRP developed by Cisco. Provides superior convergence properties and operating efficiency, and combines the advantages of link-state protocols with those of distance vector protocols.

**ESM**

Embedded Syslog Manager

**Ethernet**

Baseband LAN specification invented by Xerox Corporation and developed jointly by Xerox, Intel, and Digital Equipment Corporation. Ethernet networks use CSMA/CD and run over a variety of cable types at 10 Mbps. Ethernet is similar to the IEEE 802.3 series of standards. It is the most commonly used LAN technology, because its protocol is easy to understand, implement, manage, and maintain. It allows low-cost network implementations, provides extensive topological flexibility for network installation, and guarantees successful interconnection and operation of standards-compliant products, regardless of manufacturer.



https://t.me/learningnets

**FIB**

forwarding information base

**FTTH**

fiber to the home.

**Gigabit Ethernet**

Standard for a high-speed Ethernet, approved by the IEEE (Institute of Electrical and Electronics Engineers) 802.3z standards committee in 1996.

**GRE**

Generic Routing Encapsulation. Tunneling protocol that was developed by Cisco and that can encapsulate a variety of protocol packet types inside IP tunnels. This process creates a virtual point-to-point link to Cisco routers at remote points over an IP network.

**IANA**

Internet Assigned Numbers Authority. Organization operated under the auspices of the ISOC as a part of the IAB. IANA delegates authority for IP address-space allocation and domain-name assignment to the InterNIC and other organizations. IANA also maintains a database of assigned protocol identifiers used in the TCP/IP stack, including autonomous system numbers.

**ICMP**

Internet Control Message Protocol. Network layer Internet protocol that reports errors and provides other information relevant to IP packet processing. Documented in RFC 792.

**ICMPv6**

Internet Control Message Protocol version 6 is the implementation of Internet Control Message Protocol (ICMP) for IPv6. ICMPv5 is defined in RFC 4443.

**IETF**

Internet Engineering Task Force. Task force consisting of over 80 working groups responsible for developing Internet standards. The IETF operates under the auspices of ISOC.

**IGMP**

Internet Group Management Protocol. Used by IP hosts to report their multicast group memberships to an adjacent multicast router.

**IGMPv1**

Internet Group Management Protocol version 1

**IGMPv2**

Internet Group Management Protocol version 2

**IGMPv3**

Internet Group Management Protocol version 3

**IIF**

incoming interface

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets

**IntServ**

Integrated Services. A QoS model providing a mechanism to guarantee end-to-end QoS by explicitly managing network resources. It uses RSVP signaling for resource reservation and connection admission mechanisms to establish and maintain QoS.

**IP**

Internet Protocol. Network layer protocol in the TCP/IP stack offering a connectionless internetwork service. IP provides features for addressing, type-of-service specification, fragmentation and reassembly, and security. Defined in RFC 791.

**IP address**

32-bit address assigned to hosts using TCP/IP. An IP address belongs to one of five classes (A, B, C, D, or E) and is written as 4 octets separated by periods (dotted decimal format). Each address consists of a network number, an optional subnetwork number, and a host number. The network and subnetwork numbers together are used for routing, and the host number is used to address an individual host within the network or subnetwork. A subnet mask is used to extract network and subnetwork information from the IP address. CIDR provides a new way of representing IP addresses and subnet masks. Also called an Internet address.

**IP SLA**

(Cisco IOS) IP Service Level Agreement. Cisco IOS mechanism that provides active monitoring of synthetically generated traffic to measure performance statistics for an IP network.

**IPCP**

Internet Protocol Control Protocol. This network control protocol establishes IP over PPP.

**IPsec**

IP Security. A framework of open standards that provides data confidentiality, data integrity, and data authentication between participating peers. IPsec provides these security services at the IP layer. IPsec uses IKE to handle the negotiation of protocols and algorithms based on local policy and to generate the encryption and authentication keys to be used by IPsec. IPsec can protect one or more data flows between a pair of hosts, between a pair of security gateways, or between a security gateway and a host.

**IPv4**

IP version 4. Internet Protocol version 4 is the fourth version in the development of the IP and the first version of the protocol to be widely deployed. Along with IPv6, IPv4 is at the core of standards-based internetworking methods of the Internet. IPv4 is still used to route most traffic across the Internet. IPv4 is a connectionless protocol for use on packet-switched link layer networks (for example, Ethernet). It operates on a best-effort delivery model in that it does not guarantee delivery and does not assure proper sequencing or avoidance of duplicate delivery.

**IPv6**

IP version 6. Replacement for the current version of IP (version 4). IPv6 includes support for flow ID in the packet header, which can be used to identify flows. Formerly called IPng (next generation).

**IPv6CP**

IPv6 Control Protocol



https://t.me/learningnets

**ISP**

Internet service provider. Company that provides Internet access to other companies and individuals.

**IVR**

interactive voice response. Term used to describe systems that provide information in the form of recorded messages over telephone lines in response to user input in the form of spoken words or, more commonly, DTMF signaling. Examples include banks that allow you to check your balance from any telephone and automated stock quote systems.

**LAN**

local-area network. High-speed, low-error data network covering a relatively small geographic area (up to a few thousand meters). LANs connect workstations, peripherals, terminals, and other devices in a single building or other geographically limited area. LAN standards specify cabling and signaling at the physical and data link layers of the OSI model. Ethernet, FDDI, and Token Ring are widely used LAN technologies.

**LCP**

link control protocol. Protocol that establishes, configures, and tests data-link connections for use by PPP.

**LDP**

Label Distribution Protocol. A standard protocol between MPLS-enabled routers to negotiate the labels (addresses) used to forward packets. This protocol is not supported in Cisco IOS Release 12.0. The Cisco proprietary version of this protocol is the Tag Distribution Protocol (TDP).

**LLQ**

low latency queuing. Provides a strict priority queue mechanism to CBWFQ.

**MAC address**

Standardized data link layer address that is required for every port or device that connects to a LAN. Other devices in the network use these addresses to locate specific ports in the network and to create and update routing tables and data structures. MAC addresses are 6 bytes long and are controlled by the IEEE. Also known as a hardware address, MAC layer address, and physical address.

**MBGP**

Multicast Border Gateway Protocol.

**MD5**

Message Digest 5. A one-way hashing algorithm that produces a 128-bit hash. Both MD5 and Secure Hash Algorithm (SHA) are variations on MD4 and are designed to strengthen the security of the MD4 hashing algorithm. Cisco uses hashes for authentication within the IPsec framework. Also used for message authentication in SNMP v.2. MD5 verifies the integrity of the communication, authenticates the origin, and checks for timeliness.

**MFWD**

Multicast Forwarding Engine

**MLD**

Multicast Listener Discovery

 © 2014 Cisco Systems, Inc.

https://t.me/learningnets

**MLDv1**

Multicast Listener Discovery version 1

**MLDv2**

Multicast Listener Discovery version 2

**MP-BGP**

Multiprotocol Border Gateway Protocol.

**MRIB**

Multicast Routing Information Base

**MSDP**

Multicast Source Discovery Protocol

**multicast**

Single packets copied by the network and sent to a specific subset of network addresses. These addresses are specified in the Destination Address field.

**multicast address**

Single address that refers to multiple network devices. Synonymous with "group address."

**muRIB**

multicast-unicast RIB

**NAP**

network access point

**NAS**

network access server. Cisco platform (or collection of platforms, such as an AccessPath system) that interfaces between the packet world (for example, the Internet) and the circuit world (for example, the PSTN).

**NAT**

Network Address Translation. Mechanism for reducing the need for globally unique IP addresses. NAT allows an organization with addresses that are not globally unique to connect to the Internet by translating those addresses into globally routable address space. Also known as Network Address Translator.

**NAT44**

Network Address Translation 44. The term NAT44 is sometimes used to more specifically indicate mapping between two IPv4 addresses.

**NAT64**

Network Address Translation 64. NAT64 is a mechanism to allow IPv6 hosts to communicate with IPv4 servers.



https://t.me/learningnets

**NCP**

Network Control Protocol. Series of protocols for establishing and configuring different network layer protocols, such as for AppleTalk over PPP.

**NDP**

Neighbor Discovery Protocol.

**NSP**

network-specific prefix

**NTP**

Network Time Protocol. Protocol built on top of TCP that ensures accurate local timekeeping with reference to radio and atomic clocks located on the Internet. This protocol is capable of synchronizing distributed clocks within milliseconds over long time periods.

**NTPv4**

Network Time Protocol version 4.

**NVRAM**

nonvolatile RAM. RAM that retains its contents when a unit is powered off.

**OAM**

Operations, Administration, and Maintenance. An acronym that is used to describe the tools, processes, and standards used in the operations, administration, and maintenance of systems. Often used as a general description of these activities or to describe these activities for a specific system or protocol.

**OIL**

outgoing interface list

**OSPF**

Open Shortest Path First. Link-state, hierarchical IGP routing algorithm proposed as a successor to RIP in the Internet community. OSPF features include least-cost routing, multipath routing, and load balancing. OSPF was derived from an early version of the IS-IS protocol.

**OSPFv3**

Open Shortest Path First version 3

**PAT**

port address translation. Translation method that allows the user to conserve addresses in the global address pool by allowing source ports in TCP connections or UDP conversations to be translated. Different local addresses then map to the same global address, with port translation providing the necessary uniqueness. When translation is required, the new port number is picked out of the same range as the original following the convention of Berkeley Standard Distribution (SD).

**PE**

provider edge. Identifies the network devices, under the administrative control of the provider, that connect to CE devices.



https://t.me/learningnets

**PHB**

per-hop behavior. A description of the externally observable forwarding treatment that is applied at a DiffServ-compliant node to a behavior aggregate.

**PIM**

Protocol Independent Multicast. Multicast routing architecture that allows the addition of IP multicast routing on existing IP networks. PIM is unicast routing protocol independent and can be operated in two modes: dense and sparse.

**PIM-DM**

PIM dense mode. One of the two PIM operational modes. PIM dense mode is data-driven and resembles typical multicast routing protocols. Packets are forwarded on all outgoing interfaces until pruning and truncation occurs. In dense mode, receivers are densely populated, and it is assumed that the downstream networks want to receive and will probably use the datagrams that are forwarded to them. The cost of using dense mode is its default flooding behavior. Sometimes called dense mode PIM or PIM DM.

**PIM-SM**

PIM sparse mode. One of the two PIM operational modes. PIM sparse mode tries to constrain data distribution so that a minimal number of routers in the network receive it. Packets are sent only if they are explicitly requested at the RP (rendezvous point). In sparse mode, receivers are widely distributed, and the assumption is that downstream networks will not necessarily use the datagrams that are sent to them. The cost of using sparse mode is its reliance on the periodic refreshing of explicit join messages and its need for RPs. Sometimes called sparse mode PIM or PIM SM.

**PIM-SSM**

PIM Source-Specific Multicast

**PIMv2**

Protocol-Independent Multicast version 2

**PIMv6**

Protocol-Independent Multicast version 6

**P-network**

provider network

**POP**

point of presence. In OSS, a physical location where an interexchange carrier installed equipment to interconnect with a local exchange carrier (LEC).

**PPP**

Point-to-Point Protocol. Successor to SLIP that provides router-to-router and host-to-network connections over synchronous and asynchronous circuits. Whereas SLIP was designed to work with IP, PPP was designed to work with several network layer protocols, such as IP, IPX, and ARA. PPP also has built-in security mechanisms, such as CHAP and PAP. PPP relies on two protocols: LCP and NCP.

**PPPoE**

PPP over Ethernet.



https://t.me/learningnets

**QoS**

quality of service. Measure of performance for a transmission system that reflects its transmission quality and service availability.

**RADIUS**

Remote Authentication Dial-In User Service. Database for authenticating modem and ISDN connections and for tracking connection time.

**RFC**

Request for Comments. Document series that is used as the primary means for communicating information about the Internet. Some RFCs are designated by the IAB as Internet standards. Most RFCs document protocol specifications, such as Telnet and FTP, but some RFCs are humorous or historical. RFCs are available online from numerous sources.

**RIB**

routing information base

**RIPng**

Routing Information Protocol next generation

**RP**

rendezvous point. Router specified in PIM sparse mode implementations to track membership in multicast groups and to forward messages to known multicast group addresses.

**RPF**

Reverse Path Forwarding

**RPL**

Routing Policy Language

**RPT**

rendezvous-point tree

**RSA**

Acronym stands for Rivest, Shamir, and Adleman, the inventors of the technique. Public-key cryptographic system that can be used for encryption and authentication.

**RSVP**

Resource Reservation Protocol. A network-control protocol that allows endpoints to request specific QoS for their data flows.

**SA**

source-active

**SDR**

Session Directory



https://t.me/learningnets

**SLA**

service level agreement.

**SLA**

Site-Level Aggregator.

**SPT**

shortest path tree

**SSH**

Secure Shell Protocol. Protocol that provides a secure remote connection to a rou through a TCP application.

**SSM**

Source Specific Multicast

**TCP**

Transmission Control Protocol. Connection-oriented transport layer protocol that provides reliable full-duplex data transmission. TCP is part of the TCP/IP protocol stack.

**Telnet**

Standard terminal emulation protocol in the TCP/IP protocol stack. Telnet is used for remote terminal connection, enabling users to log into remote systems and use resources as if they were connected to a local system. Telnet is defined in RFC 854.

**TLD**

top-level domain DNS server.

**ToS**

type of service.

**TTL**

Time to Live. A mechanism that limits the lifespan or lifetime of data in a computer or network.

**UDP**

User Datagram Protocol. Connectionless transport layer protocol in the TCP/IP protocol stack. UDP is a simple protocol that exchanges datagrams without acknowledgments or guaranteed delivery, requiring that error processing and retransmission be handled by other protocols. UDP is defined in RFC 768.

**unicast**

Message sent to a single network destination.

**VLAN**

virtual LAN. Group of devices on one or more LANs that are configured (using management software) so that they can communicate as if they were attached to the same wire, when in fact they are located on a number of different LAN segments. Because VLANs are based on logical instead of physical connections, they are extremely flexible.



https://t.me/learningnets

**VPN**

virtual private network. Enables IP traffic to travel securely over a public TCP/IP network by encrypting all traffic from one network to another. A VPN uses tunneling to encrypt all information at the IP level.

**WAN**

wide-area network. Data communications network that serves users across a broad geographic area and often uses transmission devices provided by common carriers. Frame Relay, SMDS, and X.25 are examples of WANs.

**WKP**

well-known prefix

**WRED**

weighted random early detection. Queuing method that ensures that high-precedence traffic has lower loss rates than other traffic during times of congestion.

 ©2014 Cisco Systems, Inc.

https://t.me/learningnets